Studies on Intertemporal Preferences with Applications to Health Economics

Studies on Intertemporal Preferences with Applications to Health Economics

Studies naar intertemporele voorkeuren met toepassingen op gezondheidseconomie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

vrijdag 25 januari 2008 om 13.30 uur

door

Arthur Ewoud Attema

geboren te Almelo

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren:

Prof.dr. H. Bleichrodt

Prof.dr. P.P. Wakker

Overige leden:

Dr. W.B.F. Brouwer

Prof.dr. D. Read

Prof.dr.ir. A.M. Stiggelbout

Contents

1 Introduction

Economic benefits are often received at different points in time. There are numerous examples of economic applications where the outcomes occur at multiple points in time. Among these are the savings decisions of households, the environmental policies of countries, investment decisions of firms, health-related decisions of individuals, and educational activities of students.

In the majority of these cases, future outcomes are valued lower than similar present outcomes, i.e. there is positive time preference. There are several reasons for this behavior. One reason is that the future is almost always surrounded by uncertainty, whilst outcomes received immediately or in the nearer future are more certain. This translates into the discounting of future outcomes.

Second, utility is often concave in outcomes (diminishing marginal utility). This means that more units of a particular outcome give less additional utility the more one already possesses of that outcome or the more one has already consumed of it. A second cup of coffee, for example, often gives less utility than the first one. Because wealth is increasing over time due to economic growth, people have more possibilities to consume in the future than in the present. The utility of this extra consumption does, however, not increase proportionally with the increase of consumption, so that future outcomes give less utility than similar present outcomes.

Third, people tend to be myopic and do not always consider all available information about the future. This kind of behavior has the same effect as giving less weight (or no weight at all) to future outcomes.

Fourth, lifetime of individuals is finite (Bommier, 2006), whereas society has an infinite lifetime, and, hence, individuals may not care about so much about society after their lives have ended.

1.1 Measuring time preference

Time preference has profound implications for many economic choices. Therefore, it is necessary to obtain good measurements of time preference. In several scientific disciplines, including economics, psychology, and medicine, an interesting debate is going on about the proper way to discount future benefits (Frederick et al., 2002). A major part of the literature assumes *time-separability*, which means that total discounted utility can be obtained by multiplying utility in each period by a time weight and then adding up these discounted utilities. This implies that marginal utility of an outcome at some point in time is independent of the amount of that outcome at some other point in time. The most widely used discounted utility model is *constant discounting* in which the discount function is determined by a constant rate of discount. However, the practice of discounting future utility streams with a constant rate has been disputed, due to empirical violations of some axioms of the constant discounting model (e.g. Ainslie, 1975; Thaler, 1981; Benzion et al., 1989). Hyperbolic discounting models (e.g. Harvey, 1986; Loewenstein and Prelec, 1992) are popular alternatives. The discount rate is not constant but decreases with the time delay in hyperbolic discounting models, i.e. hyperbolic discounters act as if they become more patient when payoffs are more remote. Several other violations have been observed as well, including differential discounting of gains and losses (Thaler, 1981; Loewenstein, 1988).

A drawback of most of the previous empirical studies on time preference is that they assumed linear utility of money, or assumed that the utility function had a particular parametric shape. Their time preference estimates are therefore biased if this assumption does not hold. An important purpose of this thesis is to solve this problem by proposing and testing new methods for measuring time preference that do not need these assumptions. First, an intertemporal utility elicitation method is introduced that can measure utility without having to assume a particular parametric shape and subsequently can be used to correct measured time preference for utility curvature. I am able to compare different discounting models with these corrected estimates, and to indicate which model gives the best fit, which was done hardly before. I also test whether the gain-loss asymmetry can be explained by differential utility functions for gains and losses.

Another method is introduced in this thesis that allows for directly testing whether individuals deviate from the constant discounting model and to quantify their deviation from this model without having to elicit the utility function over money. The method can test whether alternative time preference models correspond better to the data. This method has a lot of potential use. It, for example, makes it possible, by means of a few simple questions, to test whether individuals are prone to intertemporal arbitrage (see Attema, 2006, for an example).

Finally, a new method is proposed to measure time preference for future life years, also known as *utility for life duration*. It is important to have knowledge about this utility function, as it is crucial in making treatment recommendations that best reflect the interests of the patient. The usual way to obtain information about this function is through the certainty equivalence method, which elicits utility under risk. This method requires *expected utility*, the normative theory for decision making under risk, to hold. Unfortunately, expected utility lacks descriptive validity (Starmer, 2000), so that the elicited utilities may be biased. In addition, these methods need the outcome death as stimulus, which tends to produce strong risk aversion and, hence, strong concavity of utility (e.g. Tversky and Kahneman, 1986; Stiggelbout and de Haes, 2001; Bleichrodt et al., 2003). It therefore seems worthwhile to find new techniques to obtain estimates of utility curvature for life duration that use a risk-free context and avoid the inclusion of the outcome death. In this thesis I propose such a technique, i.e. the *risk-free method*.

1.2 Applications of proposed measurement methods

Another purpose of this thesis is to consider a number of applications of the introduced measurement methods. First, I will investigate the universality of the utility concept. Some economists have argued that utility is only valid within the domain in which it was measured, whereas others consider utility to be a universal concept that is applicable in different contexts. I test these conjectures by comparing the results of the proposed intertemporal utility of money elicitation method to existing results generated by methods that used risky utility. In the same vein, I compare risk-free utility of life duration to risky utility of life duration.

A second application concerns the use of the risk-free method for correcting the time tradeoff (TTO) method, an important and frequently used method to measure health state utilities, for utility of life duration. I discuss each of these investigations in more detail hereafter.

1.3 Universality of utility

This thesis considers whether one unifying concept of utility exists that holds under different situations or that utility is context-dependent and varies across domains. Economists have traditionally argued that utility differs across domains and, hence, that the utility function that is relevant for decision making under risk cannot be applied in other contexts, such as decision making under certainty or intertemporal decision making (see Wakker, 1994, for an overview). In contrast, in the health economics field there is a tendency to assume transferability of utility. For example, the TTO method measures utility in an intertemporal context, but the resulting TTO utilities are often used in economic evaluations of health care, i.e. in welfare judgments. The same holds true for utilities elicited by the standard gamble method, which considers a risky situation.

This thesis experimentally measures utility functions for money and health in several decision contexts. A novelty in this thesis is that utility for money is elicited in an intertemporal setting. The results are compared to previous utility elicitations in a risky setting (Chapter 3). The risk-free method to measure the utility for life duration in a risk-free situation is proposed in Chapter 5. It is compared to the results obtained with two familiar elicitation methods that use a risky setting for the same respondents.

1.4 The time tradeoff method

The final part of this thesis applies the measurement of utility over life years to correct the TTO method for utility curvature. In a TTO, individuals need to make a tradeoff between quality of life and duration of life. A problem of the TTO method is, however, that it assumes linear utility of life duration, whereas this is often found to be concave, because many people discount future lifetime. This results in a downward bias of health state utilities (Bleichrodt, 2002). It is desirable to quantify this bias and to correct for it.

There have been done some previous attempts to correct TTO scores for the utility of life duration (e.g. Stiggelbout et al., 1994; Stalmeier et al., 1996; van Osch et al., 2004; van der Pol and Roux, 2005), but most of these studies used the CE method and therefore required expected utility to hold. When expected utility does not hold, the correction of TTO scores will be biased. In this thesis the risk-free method is employed to correct TTO scores for utility of life duration curvature, so that one is not dependent on the validity of expected utility and the influence of the outcome death. The differences with uncorrected TTO scores are investigated and the role of utility correction in several violations of the TTO method is explored.

1.5 Outline

The structure of the thesis is as follows. It begins with an overview of the available evidence on time preference in Chapter 2, with special attention to the health economics field. The most important findings are presented and their implications for medical decision making are discussed. Chapters 3 and 4 continue with experiments concerning the measurement of time preference. These chapters are of a general nature and therefore not specific to health outcomes. Chapter 3 develops a new method to measure time preference that corrects for utility curvature in a nonparametric way. This method is subsequently employed in an experiment to estimate utility and time preference for both gains and losses. Chapter 4 introduces another method that enables us to quantify the deviation from constant discounting without having to elicit the utility function over money. An experiment is presented to test this method and its results are discussed.

In the remainder of this thesis, I focus on the health domain. Chapter 5 proposes and tests the risk-free method to measure the utility function for life duration. In addition, it compares this method with two risky methods and presents the results of a questionnaire about the feasibility of these methods. Chapter 6, 7, and 8 use the new method of Chapter 5 in TTO measurements. In Chapter 6 it is explained how the risk-free method can be used to correct TTO scores for utility of life duration curvature. I measure the magnitude of this correction by means of an experiment. Chapter 7 investigates whether the elicitation procedure used in the TTO method influences its results and whether this influence is diminished when correcting for utility of life duration. Chapter 8 tests an important assumption of the TTO method, known as the assumption of constant proportional tradeoffs (CPTOs). This means that individuals are willing to give up the same proportion of lifetime irrespective of its duration. The empirical evidence about this assumption so far available is reviewed and I test whether individuals also (or instead) constantly proportionally trade off *utilities*. Finally, Chapter 9 discusses the main findings of this thesis and concludes.

2 Developments in time preference and their implications for medical decision making

Summary

The field of time preference is developing rapidly. It concerns important concepts for many economic issues. One important domain of application is health economics. This chapter reviews several empirical and theoretical developments for time preference with special attention to applications in health economics.

2.1 Introduction

Costs and benefits often occur at different points in time. This concerns concepts of time preference, because costs and benefits in the future are often valued less than when they occur immediately (see Olson and Bailey (1981) for arguments in favor of positive discounting).

The last decades have shown some interesting developments in the area of time preference. We review the literature to date and in particular we consider the impact on the health economics field. Time preference is an important factor in health economics for at least three reasons. First, it plays a role in several methods that measure individual preferences for health states. For example, the time tradeoff (TTO) method, one of the most popular health valuation methods, is heavily dependent upon time preference. Individuals may incorporate their time preferences in the assessment process, which may cause double discounting of QALYs when the elicited TTO scores are discounted by a standard discount rate (e.g. MacKeigan et al., 2003; Gravelle et al., 2007). In addition, individuals may have different discount structures, so that these need to be measured first in order to obtain a robust estimate of a health state (e.g. Bleichrodt, 2002; Attema and Brouwer, 2007a).

Second, time preference has an impact on health-affecting behavior. This is because this kind of behavior often has short-term costs and long-term benefits or vice versa. For example, smoking involves immediate benefits (e.g. relief from craving, physical reactions, and so on) whereas most of the costs are not visible immediately (e.g. lung cancer or a heart attack in the future). Likewise, time preference can influence many health-related activities, like exercising, dieting, several addictive habits, and performing dangerous jobs. The relation between time preference and health behavior is not very clear though. Contrary to expectations, the link between discounting and preventive health behavior appears to be very weak (e.g. Fuchs, 1982; Chapman and Coups, 1999; Chapman et al., 2001) or not to exist at all (e.g. Chapman, 1998). Only for addictive behavior there seems to be a substantial relation with time preference (e.g. Vuchinich and Simpson, 1998; Bickel et al., 1999; Bretteville-Jensen, 1999; Kirby et al., 1999; Madden et al., 1999; Baker et al., 2003). Chapman (2005) presents a metaanalysis and finds no significant correlation between time preference and health behavior, but she does find a significant correlation between time preference and addictive behavior. However, in the case of addiction, it is difficult to speculate on the direction of causation, as it may well be that being addicted causes discounting to increase instead of the reverse. Another question is whether the methods that elicited time preferences in these studies really capture the degree of impatience that is relevant for health purposes. More research in time preference is therefore necessary in order to get a better understanding of the relation between time preference and health-affecting behavior.

Third, public policy makers frequently make decisions about future health outcomes. These decisions require economic evaluations of health care programs, in which time preference is a necessary input.

This chapter is organized in the following way. Section 2.2 presents the most important theories that have been proposed to model intertemporal choice. Section 2.3 describes the empirical evidence, in particular the violations of the classical theory that were found. Finally, Section 2.4 discusses the implications of these findings for medical decision making.

2.2 Review of theoretical developments

Time preference was first formally described by Ramsey (1928), and Fisher (1930) was the first to use an indifference framework for analyzing the discount rate. However, the paper by Samuelson (1937) has become the classical study on time preference. He proposed a formal way to model time preference, which is known as *constant* or *exponential discounting*:

$$V(x_0,\ldots,x_T)=\sum_{t=0}^{T}\delta^t u(x_t), \tag{1}$$

with V representing utility over a stream of outcomes x over time t, T the final period of the considered horizon, δ the discount factor, and u a real-valued instantaneous utility function that represents preferences over outcomes. Koopmans (1960) axiomatized this model, which has become the standard discounted utility model in economics.

A lot of studies have cast doubts on the descriptive validity of the constant discounting model, however, as will be shown in the next section. To accommodate the observed violations, alternative discounting models have been proposed. Loewenstein and Prelec (1992) introduced the generalized hyperbolic discounting function:

$$V(x_0,\ldots,x_T)=\sum_{t=0}^{T}\left(\frac{1}{1+gt}\right)^{h/g}u(x_t), \tag{2}$$

with g,h > 0 and g determining the departure from constant discounting, while h indicates the magnitude of discounting. Constant discounting is the limiting case of this function for g→0. A positive value of g implies that discount rates decline over time instead of staying constant. We will call this phenomenon *decreasing impatience* throughout the paper. Similarly, increasing discount rates will be described by *increasing impatience*. The generalized hyperbolic discounting function was applied by Green et al. (1994), Cairs and van der Pol (1997a, 2000), Antonides and Wunderink (2001), and van der Pol and Cairns (2002), among others.

Herrnstein (1981) proposed a special case of the generalized hyperbolic discounting model where h = g, which was axiomatized by Harvey (1994). This model is very popular among psychologists (e.g. Mazur, 1987; Rachlin et al., 1991; Kirby and Marakovic, 1995; Myerson and Green, 1995; Kirby et al. 1999; Green et al., 2005). Harvey (1986) gave an axiomatic foundation for another special case of the model of Loewenstein and Prelec (1992), with h = 1. Cropper et al. (1994) used this model when analyzing the discounting of future lives saved and van der Pol and Cairns (2002) applied it to non-fatal changes in health.

The most popular hyperbolic discounting model among economists is the quasi-hyperbolic discounting model (Phelps and Pollak, 1968). This model can be represented as follows:

$$V(x_0, \ldots, x_T) = u(x_0) + \sum_{t=1}^{T} \beta\delta^t u(x_t), \qquad (3)$$

with $0 < \beta \leq 1$. The only difference with constant discounting is the parameter β. Constant discounting is the special case of quasi-hyperbolic discounting with β = 1. If $\beta < 1$ then the outcome in the first period is discounted at a higher rate than the discount rate that is used to compare the outcomes in any two other contiguous future periods. In other words, $\beta < 1$ models an *immediacy effect.* Quasi-hyperbolic discounting was popularized in economic applications by

Laibson (1997). Its advantage is that it needs only one parameter, it is analytically tractable, and it still captures the essential characteristic of hyperbolic discounting, i.e. decreasing impatience.

Quasi-hyperbolic discounting has been used to explain several economic anomalies, including health-affecting behavior (e.g. Brocas and Carrillo, 2001; Gruber and Köszegi, 2001; Harris and Laibson, 2001; Bénabou and Tirole, 2002; Diamond and Köszegi, 2003; Krusell and Smith, 2003; DellaVigna and Malmendier, 2006).

Discounting of health outcomes

Health is a unique commodity, which, in contrast to money and most other commodities, cannot be transferred across time or individuals and is irreversible in nature. In addition, market forces are not very prominent in health care, so that it is difficult to measure time preferences for health from field studies. Another problem in the measurement of time preferences for health outcomes is that health states have a duration inextricably bound to it, whereas monetary amounts can be delivered at a single point in time (Dolan and Gudex, 1995; Gafni, 1995; Bleichrodt and Johannesson, 2001; Chapman, 2003).

It is questionable whether the constant discounting model should be viewed as the normatively desirable model in the health domain. Bleichrodt and Gafni (1996) gave arguments against the use of the constant discounting model for prescriptive purposes in medical decision making. Further, as will be shown in the next section, the descriptive validity of the constant discounting model in health economics is even more doubtful.

2.3 Review of empirical developments

Decreasing impatience

Probably the most robust violation of the traditional constant discounting model entails the observation that individuals often tend to discount at a decreasing instead of constant rate over time (e.g. Thaler, 1981; Benzion et al., 1989; Rachlin et al., 1991; Kirby and Herrnstein, 1995; Kirby and Marakovic, 1995; Myerson and Green, 1995; Kirby, 1997; Green et al., 2005). Kirby and Santiesteban (2003) found this pattern even when taking into account the presence of concave utility, transaction costs and risk. Abdellaoui et al. (2006) found it when correcting for utility curvature and without possible bias due to transaction costs and risk. The same pattern was observed for health outcomes (e.g. Loewenstein and Thaler, 1989; Cropper et al., 1992, 1994; Redelmeier and Heller, 1993; Cairns, 1994; Chapman, 1996a; Cairns and van der Pol, 1997b; Bleichrodt and Johannesson, 2001; Lazaro et al., 2001; van der Pol and Cairns, 2002).

On the other hand, there are also studies arguing against hyperbolic discounting. For example, Read (2001) suggests that the observed decreasing discount rates are caused by subadditive instead of hyperbolic discounting. That is, people tend to discount a time interval more heavily when it is divided into several parts than when considered completely. Harrison et al. (2002) find no evidence against constant discounting in an experimental study with a representative sample of the population of Denmark using real incentives. Rubinstein (2003) presents experimental results rejecting hyperbolic discounting and proposes similarity relations to account for observed intertemporal choices. Coller et al. (2005) find that subjects are divided roughly equally between those following the constant discounting model and those following the quasi-hyperbolic discounting model. The results of Sopher and Sheth (2006) can only for a small proportion be explained by hyperbolic discounting. Benhabib et al. (2006) find evidence of a small present bias in the form of a fixed cost and no rejection of constant discounting when taking this into account.

Increasing impatience

In spite of the growing interest in hyperbolic discounting models, there are also studies that find increasing instead of decreasing impatience. Examples for monetary outcomes include Gigliotti and Sopher (2003), Read et al. (2005a), Attema et al. (2006) and Sayman and Onculer (2006).

In the health economics literature there is also some, albeit indirect, evidence of increasing impatience. Martin et al. (2000) and Attema et al. (2007) estimated the utility function for life duration and found increasing absolute risk aversion over time. When we keep in mind that the utility of life duration in effect captures the rate of time preference for life years, this finding is equivalent to an increasing discount rate over time.

Sign effect

Another stylized fact in intertemporal choice is the finding of lower discounting of losses than gains of a similar magnitude. Several studies found individuals to make decisions from some reference point and to treat gains and losses seen from this reference point differently (e.g. Thaler, 1981; Loewenstein, 1988; Benzion et al., 1989; Hesketh, 2000; Antonides and Wunderink, 2001; Donkers et al., 2004). Ahlbrecht and Weber (1997), on the other hand, find no significant differences between discount rates for losses and those for gains. In the health domain, a gain-loss asymmetry was observed by MacKeigan et al. (1993) and Chapman (1996a). Several explanations for this behavior have been put forward. Loewenstein and Prelec (1992) proposed different utility functions for gains and losses. Shelley (1993) argued that losses were discounted differently merely because of a framing effect in previous studies. Both of these arguments were rejected, however, by the results of Abdellaoui et al. (2006), who found a significant sign effect even when correcting for differential utility functions for gains and losses and using a neutral frame.

Sequence effect

An anomaly of the constant discounting model that was found both for monetary and health outcomes, concerns the preference of individuals for increasing sequences over time (e.g. Hsee et al., 1991; Loewenstein and Prelec, 1991; Loewenstein and Sicherman, 1991; Varey and Kahneman, 1992; Loewenstein and Prelec, 1993). An explanation for this finding may be that individuals compare an outcome in a sequence to the outcome in the period before it, and therefore interpret a declining sequence as a series of losses. Due to loss aversion (Tversky and Kahneman, 1991), these are penalized and declining sequences get less overall utility than improving sequences, even when the former have higher net present values for any positive discount rate than the latter. In other words, it is equivalent to a negative time preference rate, even though most of these people discount future amounts positively in other situations. Chapman (1996b) suggests that decision makers compare a sequence to a reference sequence, i.e. to their expectations about how outcomes usually change over time.

Chapman (2000) finds that although sequences generally elicit a preference for improvement, this pattern is moderated by expectations about how outcomes are generally experienced over time. When the expectation of decline is strong enough, as in many aspects of health, a preference for a worsening sequence appears.

Magnitude effect

Another empirical regularity concerns the finding of lower discount rates for higher amounts than for lower amounts. This has been found both for studies with hypothetical monetary incentives (e.g. Thaler, 1981; Benzion et al., 1989; Read, 2001) and for studies with real monetary incentives (e.g. Holcomb and Nelson, 1992; Green et al., 1994; Kirby et al., 1999; Antonides and Wunderink, 2001). In addition, a similar effect was found for health outcomes (e.g. Chapman and Elstein, 1995; Chapman, 1996a; Bickel and Johnson, 2003). An exception includes the study of Gigliotti and Sopher (2003).

Most of these studies may have found a spurious magnitude effect, however, as they did not correct for utility curvature. Loewenstein and Prelec (1992) proposed a possible explanation for the magnitude effect. They attribute it to the elasticity of the utility function. Chapman (2003) concludes that the magnitude effect is solely caused by utility curvature, but this conclusion seems premature, as the empirical evidence concerning the amount of utility curvature suggests that incorrectly assuming linear utility cannot fully account for the magnitude effect. For example, economists often assume a power utility function over money (Wakker, 2007), which has the property of constant relative risk aversion, but this cannot account for a magnitude effect. Suppose a decision maker has a power utility function over money of the form $u = \sqrt{x}$ and is indifferent between 100 Euro now and 225 Euro in one year. This means that his utility discount factor is $\frac{\sqrt{100}}{\sqrt{225}} = \frac{6}{9}$ whereas not correcting for utility curvature gives a discount factor of $\frac{100}{225} = \frac{4}{9}$. If we then multiply the monetary amount that can be received immediately by a factor 2, the decision maker can get a utility of $\sqrt{200} \approx 14.14$ right now, which is equivalent to a utility of $\frac{14.14}{6/9} \approx 21.21$ in one year. Therefore, the monetary amount that makes him indifferent to 200 Euro immediately is $21.21^2 = 450$ in one year. Not correcting for utility curvature again gives a discount factor of $\frac{4}{9}$, so that there is no magnitude effect even when incorrectly assuming linear utility. Consequently, one has to assume a utility function that does not belong to the family of constant relative risk aversion. An alternative is the exponential utility function, which belongs to the constant *absolute* risk aversion family and which does predict a magnitude effect when incorrectly assuming linear utility. However, the amount of the magnitude effect that is found mostly is so high (e.g. Thaler, 1981) that it would require absurdly high utility curvature to have discount functions that are independent of the magnitude of the outcomes.

Framing

The existing empirical literature on time preference demonstrates the considerable role the experimental design can play. Potentially influential factors include the response mode (money or time, see for example Attema et al., 2006), the framing of questions (e.g. Shelley, 1993), the time unit used (e.g. days, weeks, months or years), and the elicitation procedure (choices versus matching, e.g. Ahlbrecht and Weber, 1997).

Another framing effect is the dependence of discounting rates on the framing of a question as the delay or speeding up of some receipt or payment (Loewenstein, 1988; Shelley, 1993; Shelley and Omer, 1996). This phenomenon seems to be related to the WTP-WTA gap and can to a large extent be explained by loss aversion (e.g. Kahneman et al., 1990). There are also several studies that found higher discounting when the same amount of delay was expressed in terms of duration than when it was expressed in terms of dates (e.g. Read et al., 2005b; LeBoeuf, 2006).

Domain independence

A further question concerns the existence of a unifying concept of discounting for several domains. Some studies have compared the discounting of monetary and health outcomes, with mixed results. Moore and Viscusi (1990) and Cropper et al. (1994) found discount rates for money and health to be similar. Chapman and Elstein (1995) reported low correlation in discount rates between money and health. Cairns (1992) found higher discount rates for money than for health. Chapman (1996a) observed domain independence even when expressing money and health on a common utility scale. Chapman et al. (1999) investigated whether preferences could be explained by familiarity with the health state and its expected development during lifetime. It appeared to play no role, however, and therefore familiarity also seems not able to explain the independence between the monetary and the health domain. In addition, dissimilarities in the descriptions of health and money could not explain the differences.

2.4 Discussion and implications for medical decision making

The evidence presented above lends support to a tendency to discount future outcomes in a hyperbolic way, both for money and for health. This has important implications for the understanding and prediction of health-affecting behavior, for the measurement of health-related preferences, and for economic evaluations. For example, when people, having a hyperbolic discount structure, give relatively more weight to the far future than under constant discounting, this implies more attention will be given to preventive programs, which generally have benefits occurring long after implementation of the program (e.g. vaccination programs). On the other hand, hyperbolic discounting may cause self-control problems (O'Donoghue and Rabin, 1999), leading individuals to repeatedly postpone investments in health, like exercising or quitting smoking.

The findings of decreasing discount rates over time, lower discounting of losses than of gains, a lower discounting of high outcomes, and a general preference for improving sequences, seem to be robust across different domains. This suggests there is some underlying mechanism that is valid for both monetary and health outcomes. On the other hand, the correlation between time preference rates for money and health is rather low, indicating that several other contextual factors are also influential in determining the observed intertemporal behavior.

Implications for health-affecting behavior

Hyperbolic discounting can be used to explain a variety of observed health actions. For example, Gruber and Köszegi (2001) have incorporated the quasi-hyperbolic discounting model into the rational addiction model of Becker and Murphy (1988). Their analysis shows that the presence of hyperbolic discounting can be a reason to introduce a tax to improve long-term welfare even when there are no externalities or market power present. As a test of this hypothesis, Gruber and Mullainathan (2005) analyzed by means of the results of a questionnaire whether higher cigarette taxes make consumers happier because it encourages them to stop smoking, which they wanted to anyway but were not able to due to self-control problems. They find that this is indeed the case, which suggests that increasing taxation on smoking may be welfare-improving.

O'Donoghue and Rabin (2006) show that taxing sin goods can increase consumer surplus when only a fraction of the population behaves time-inconsistently. These consumers have self-control problems and consume more of some unhealthy product than they actually want. Redistributing the obtained tax proceeds to consumers may even yield a Pareto-improving outcome where nobody is worse off. Moreover, the resulting decrease in consumption will improve society's health.

DellaVigna and Malmendier (2006) mention hyperbolic discounting as a potential explanation for their finding that consumers buy monthly contracts at gyms that turn out to be more expensive per visit than it would be if they bought a separate ticket for each visit. They probably had too high expectations about their expected frequency of visits to the gym and according to those expectations the long-term package turned out to be the most favorable. However, because of self-control problems, these expectations could not be made true and they ended up visiting the gym much less than expected. Therefore the ex post costs per visit were higher than the costs of a one visit ticket.

Contrary to hyperbolic discounting, the applications of the other anomalies to the health context are scarce. Since the evidence of a sign effect, a magnitude effect and framing is exhaustive, and relevant for medical decision making, there is need for future research to fill this gap. The gain-loss asymmetry implies that people will discount health outcomes that are framed as losses as seen from a certain reference point less than when these are framed as gains. This can be used by policy makers in order to achieve goals regarding better long-term health. For example, when governments want to reduce smoking in a country, it may help to focus on the future losses that will be incurred when one continues to smoke, instead of focusing on the future gains in terms of better health when one stops smoking. An implication of the magnitude effect is that emphasizing outcomes of large magnitudes can encourage future-oriented behavior, as these will be discounted at a lower rate.

Implications for the measurement of health preferences

The evidence on time preference also has implications for health valuation methods. The findings against constant relative risk aversion in the context of life duration are negative evidence for the QALY model as proposed by Pliskin et al. (1980). If their assumptions hold, it necessarily follows that the utility function for life years can be represented by a power function, which implies constant relative risk aversion. In addition, its implied condition of constant proportional tradeoffs has often been violated (e.g. Stiggelbout et al., 1995; Bleichrodt et al., 2003; Attema and Brouwer, 2007c).

The evidence against constant discounting has implications for the QALY model as well, since future QALYs are often discounted at a constant rate. Economic evaluations may still incorporate constantly discounted QALYs for normative purposes, however, as long as the elicited quality weights are not distorted by time preference of respondents incorporated into their answers. For instance, raw TTO scores may already reflect time preference so that discounting them would underestimate utility.

3 Intertemporal tradeoffs for gains and losses: An experimental measurement of discounted utility[1]

Summary

This chapter is the first to measure utility in intertemporal choice and presents new and more robust evidence on the discounting of money outcomes. Our measurement method is parameter-free in the sense that it requires no assumptions about utility or discounting. We found that intertemporal utility was concave for gains and convex for losses, consistent with a hypothesis put forward by Loewenstein and Prelec (1992). Utility in intertemporal choice was close to utility in decision under risk and uncertainty, suggesting that there may be one unifying concept of utility that applies to all of economics. The existence of one concept of utility is important for applied economics, because it largely reduces data requirements. Discount rates declined over time, but less so than has been observed in previous studies that assumed linear utility. Of the main discounted utility models, Loewenstein and Prelec's (1992) generalized hyperbolic

[1] This chapter is based on Abdellaoui, Attema, and Bleichrodt (2006).

discounting model best fitted our data. The widely used quasi-hyperbolic model fitted the data only slightly better than constant discounting. Finally, we obtained evidence of an asymmetry in discounting between gains and losses, which, in contrast with earlier findings, cannot be explained by a framing effect.

3.1 Introduction

Many economic decisions involve outcomes that occur at different points in time. To model such decisions, discounted utility models are typically used. These models combine a utility function that reflects attitudes towards outcomes and a discount function that captures the effect of the passage of time. The most widely used discounted utility model in economics is constant discounting in which the discount function is determined by a constant rate of discount. Empirical studies on time preference have observed that discount rates are not constant but decrease over time, a phenomenon referred to as decreasing impatience (Frederick et al., 2002; Read, 2004). These findings have led to the development of alternative discounted utility models, commonly referred to as hyperbolic discounting. The hyperbolic discounting models are consistent with decreasing impatience and have become quickly popular in economics. Today many applications are based on hyperbolic discounting, in particular on quasi-hyperbolic discounting, a model that was first proposed by Phelps and Polak (1968) and made popular by Laibson (1997).[2]

Empirical measurement of discounted utility models is complex, because it requires the simultaneous elicitation of the utility function and the discount function. Previous studies have side-stepped this problem and have assumed specific functional forms for the utility function and the discount function. In particular, most studies have assumed linear utility. A drawback of making parametric assumptions is that the quality of the estimation comes to depend on the choice of functional forms. For example, if utility is concave instead of linear then falsely assuming linear utility will lead to an overestimation of discount rates. It should be noted that most empirical studies have indeed found high discount rates. Another limitation of assuming functional forms for utility is that no or only limited information is obtained on the intertemporal utility function. Consequently, in spite of the importance of intertemporal preferences and discounted utility models in economics, there exists to date no study that has actually measured the utility function in intertemporal choice.

In light of the above problems, this chapter presents a new method to measure both intertemporal utility and the discount function without making any assumptions about functional forms. It is in this sense that we refer to our method as parameter-free. An additional advantage of our method is that it allows

---

[2] Examples of applications based on quasi-hyperbolic discounting include Laibson (1997), Bernheim et al. (2001), Harris and Laibson (2001), Krusell and Smith (2003), and Salanié and Treich (2006) for saving, O'Donoghue and Rabin (1999) and Brocas and Carrillo (2001) for procrastination, Brocas and Carrillo (2000) for the value of information, Gruber and Köszegi (2001) for addiction, Bénabou and Tirole (2002) for self-confidence, Diamond and Köszegi (2003) for retirement, and Karp (2005) for global warming.

measuring utility and discounting at the individual level and, therefore, takes account of heterogeneity in individual preferences. We applied our method in an experimental study and, hence, this paper is the first to measure intertemporal utility and to obtain robust estimates of the discount function at the individual level.

Our data allowed us to address several empirical questions. First, we obtained evidence on the shape of the utility function in intertemporal choice. Classical economics assumes that this utility function is everywhere concave. Loewenstein and Prelec (1992), by contrast, posit that people treat gains and losses differently and have concave utility for gains and convex utility for losses. We performed our experiment both for gains and for losses, which made it possible to compare the predictions of classical economics with Loewenstein and Prelec's (1992) hypotheses.

Second, our findings on intertemporal utility allowed us to shed some empirical light on a long-standing issue in economics, whether there exists one unifying concept of utility that applies to all of economics or whether different concepts of utility apply in different decision contexts (for a history see Wakker, 1994). No empirical guidance exists on this debate, because few measurements of utility were available in decision contexts other than decision under risk. The existence of one unifying concept of utility would be highly desirable for applied economics because it implies that utility measurements can be transferred across decision contexts thereby greatly reducing data requirements. By comparing our utility measurements with those from the literature on decision under risk, we were able to provide insights regarding the existence of one unifying concept of utility.

Third, we could test whether the commonly observed pattern of declining discount rates persisted when the assumption of linear utility was relaxed. As argued above, there are grounds to suspect that previous observations of decreasing impatience may, at least partly, have been caused by falsely assuming linear utility. Our data also made it possible to compare the fit of constant discounting with that of its main hyperbolic alternative, quasi-hyperbolic discounting, and three other popular hyperbolic discounting models. Many studies have provided support for hyperbolic discounting (e.g. Ainslie, 1975; Thaler, 1981; Benzion et al., 1989; Kirby and Marakovic, 1995),[3] but little insight exists into which hyperbolic model most accurately describes intertemporal preferences. The popularity of quasi-hyperbolic discounting relative to other hyperbolic discounting models in economics is based on its theoretical tractability and not on its displayed descriptive superiority.

Finally, we could perform a robust test for an asymmetry between the discounting of gains and losses that has been observed in some earlier studies (Thaler, 1981; Benzion et al., 1989). One explanation for the gain-loss asymmetry may be that it is an artifact of the assumption of linear utility. When utility is concave for gains, leading to an overestimation of discount rates for gains, and closer to linear for losses, leading to less distortion of discount rates for losses, then the gain-loss asymmetry will follow from the assumption of linear utility even when people have the same discount function for gains and for losses. The pattern "concave utility for gains and more linear utility for losses" has been observed in several empirical studies on decision under risk (Fishburn and Kochenberger, 1979; Abdellaoui, 2000; Pennings and Smidts, 2003).

[3] For findings challenging hyperbolic discounting see Read (2001) and Read et al. (2005b).

In what follows, Section 3.2 reviews previous theoretical and empirical research on intertemporal choice. Section 3.3 presents our method for measuring discounted utility. The design and results of our experiment are specified in Sections 3.4 and 3.5 and are discussed in Section 3.6. We conclude in Section 3.7.

3.2 Theory and existing empirical evidence

We consider *temporal profiles* $(x_0,\ldots,x_T)$, where x_t denotes outcome x at time point t and time point 0 is the present. Outcomes can be money amounts but also binary prospects (p:M; m) denoting money amount M with probability p and money amount m with probability 1−p. Throughout we assume that 0, i.e. receiving nothing, belongs to the set of outcomes.

We examine preferences $\succcurlyeq$ over temporal profiles. The relations $\preccurlyeq$, $\prec$, $\succ$, and $\sim$ are as usual. Preferences over outcomes are derived from preferences over *constant temporal profiles*, where $x_1 = \ldots = x_T = x$. We say that $\alpha \succcurlyeq \beta$ if and only if $(\alpha,\ldots,\alpha) \succcurlyeq (\beta,\ldots,\beta)$, i.e. receiving α at all points in time is preferred to receiving β at all points in time.

We assume that the decision maker perceives outcomes relative to 0. *Gains* are outcomes preferred to 0 and *losses* are outcomes less preferred than 0. We will only consider temporal prospects where all outcomes have the same sign, i.e. either all outcomes are gains or all outcomes are losses. A function V *represents* $\succcurlyeq$ when for all x,y, $x \succcurlyeq y$ if and only if $V(x) \geq V(y)$. Throughout, we will assume that preferences over temporal profiles can be represented by the *general discounting* model

$$V(x_0,\ldots,x_T) = \sum_{t=0}^{T} \lambda_t^i u(x_t),\ i = +,- \qquad (1)$$

with the *time weights* λ_t^i positive and $\lambda_0^i = 1$ and u a real-valued *utility function* that represents preferences over outcomes. We allow that the time weights differ for gains and for losses. To keep the notation tractable, we will suppress the sign-dependence of the λ_t^i and simply write λ_t in what follows. Whether the time weights for gains or the time weights for losses apply will be apparent from the decision context.

The time weights λ_t are unique and the utility function is unique up to unit. Equation (1) is general in the sense that it presumes nothing about the ordering or the relative magnitude of the λ_t. The main models of discounting are all special cases of (1). A preference foundation for general discounting has been given by Krantz et al. (1971, Theorem 6.15).

The best-known special case of (1) is constant discounting, which was introduced by Samuelson (1937) and which is still the most widely used discounted utility model in economics. *Constant discounting* entails that the time weights λ_t in (1) are equal to $\frac{1}{(1+\rho)^t}$, where ρ is the constant discount rate. As mentioned before, experimental evidence has cast doubts on the descriptive validity of the constant discounting. In this chapter we focus on two violations of constant discounting: *decreasing impatience,* the finding that discount rates are not constant but decrease over time, and the *gain-loss asymmetry*, the finding that people discount gains and losses differently.[4]

Many studies have observed decreasing impatience. See for example Thaler (1981), Benzion et al. (1989), Shelley (1993), and Kirby and Marakovic (1995) for money amounts and Chapman (1996a), Lazaro et al. (2001), and van der Pol and Cairns (2002) for health. An exception is Ahlbrecht and Weber (1997), who only observed decreasing impatience in a matching task, but not in a choice task. The common assumption in these studies was linear intertemporal utility. Chapman (1996a) also considered power utility. She elicited utility in an atemporal setting using introspective strength of preference judgments and then assumed that this function could also be applied to intertemporal choice.[5] Whether utility is transferable across decision domains is highly controversial in economics. Arrow (1951), Savage (1954), Luce and Raiffa (1957), and Fishburn (1989) amongst others have argued against such transferability.

There is some controversy in the literature as to whether decreasing impatience holds in general or whether violations of constant discounting occur only in the first time interval. The latter hypothesis is referred to as the *immediacy effect* and underlies quasi-hyperbolic discounting, which will be discussed below. Some studies found support for the immediacy effect (Bleichrodt and Johannesson, 2001; Frederick et al., 2002); others rejected it and also found

---

[4] The gain-loss asymmetry can be accommodated within the general discounting model because we allow that the λ_t differ between gains and losses.

[5] In a recent working paper Andersen et al. (2006) used a comparable strategy: they estimated power utility from decision under risk and then applied this function to intertemporal choice.

violations of constant discounting for later time intervals (Kirby and Herrnstein, 1995; Kirby, 1997; Lazaro et al., 2001).

The gain-loss asymmetry is empirically less well-established than decreasing impatience. Thaler (1981) and Benzion et al. (1989) found evidence of the gain-loss asymmetry, but Shelley (1993) showed that their findings could be explained by a framing effect. In a neutral frame, she found no evidence of a gain-loss asymmetry, a finding that was later confirmed by Ahlbrecht and Weber (1997).

3.2.1 Alternative discounting models

Several alternative discounting models have been proposed in response to the observed violations of constant discounting. These models were primarily designed to explain decreasing impatience. Except for the model of Loewenstein and Prelec (1992), they make no distinction between gains and losses and, hence, cannot explain the gain-loss asymmetry.

Loewenstein and Prelec (1992) suggested to use a *generalized hyperbolic discounting* model, in which $\lambda_t = \frac{1}{(1+\gamma t)^{\alpha/\gamma}}$, with $\alpha,\gamma > 0$. The parameter γ determines the departure from constant discounting. The limiting case of γ tending to zero yields constant discounting. Because α is positive, the discount rates implied by generalized hyperbolic discounting decrease over time, corresponding to decreasing impatience. Loewenstein and Prelec assumed that the time weights were the same for gains and for losses. To explain the gain-loss asymmetry, they suggested that the intertemporal utility function u in (1) is concave for gains and convex for losses and is more elastic for losses than for gains. The special case of generalized hyperbolic discounting in which $\alpha = \gamma$ was initially proposed by Herrnstein (1981) and is referred to as *proportional* or *hyperbolic* discounting. *Power discounting* (Harvey, 1986), is the special case of generalized hyperbolic discounting in which $\gamma = 1$.

Quasi-hyperbolic discounting (Phelps and Pollak, 1968; Laibson, 1997), is the special case of (1) where $\lambda_t = \dfrac{\beta}{(1+\rho)^t}$ for $t > 0$ with $0 < \beta \leq 1$. The only difference with constant discounting is the parameter β. Constant discounting is the special case of quasi-hyperbolic discounting with $\beta = 1$. If $\beta < 1$ then the outcome in the first period is discounted at a higher rate than the discount rate that is used to compare the outcomes in any two other contiguous future periods. In other words, $\beta < 1$ models the immediacy effect.

3.3 Measurement method

Our method to measure the general discounting model (1) consisted of two stages. In the first stage, choices between temporal profiles were constructed in such a manner that the time weights λ_t canceled, allowing us to measure utility without the need to know the time weights. This way of measuring utilities resembles the utility measurement method of Wakker and Deneffe (1996) for decision under uncertainty. The difficulty in translating their method to intertemporal choice is that the utility function in intertemporal choice has different uniqueness properties than the utility function in decision under uncertainty. In the second stage, we used the elicited utilities to measure the time weights. Hence, we could measure the time weights from the elicited utilities and no assumptions about the shape of utility had to be made.

Our method involves choosing the utility of two outcomes and, as is shown in Appendix 3A, this is only allowed when all temporal profiles involve the same unit of time and have common final periods. That is, for all profiles $(x_1,\ldots,x_T)$ and $(y_1,\ldots,y_S)$, the difference in timing between x_{t-1} and x_t is equal to the difference in timing between y_{t-1} and y_t and $T = S$. Hence, we will assume such profiles in the remainder of this section and we only used such profiles in our experimental study described in Section 3.4.

First stage: measurement of utility

Let x_0y_t denote the temporal profile that gives x now, y at time point t and nothing in all other periods. The first step in the measurement of utility was to select two *gauge outcomes* G and g and a *starting outcome* x^0. Superscripts serve to distinguish outcomes and do not denote powers. We then elicited the outcome x^1 such that a participant was indifferent between $g_0x^1_t$ and $G_0\,x^0_t$. In terms of the general discounting model (1) this indifference implies that

$$u(g) + \lambda_t u(x^1) + \sum_{s\neq 0,t} \lambda_s u(0) = u(G) + \lambda_t u(x^0) + \sum_{s\neq 0,t} \lambda_s u(0) \tag{2}$$

and, hence,

$$u(x^1) - u(x^0) = \frac{u(G)-u(g)}{\lambda_t}\,. \tag{3}$$

The outcome x^1 was used as an input in the next question where we elicited x^2 such that indifference held between $g_0x_t^2$ and $G_0x_t^1$. By (1) and a similar argument as above, this indifference implied that

$$u(x^2) - u(x^1) = \frac{u(G)-u(g)}{\lambda_t}. \tag{4}$$

Thus, $u(x^2) - u(x^1) = u(x^1) - u(x^0)$. We proceeded to elicit indifferences $g_0x_t^j \sim G_0x_t^{j-1}$, $j = 3,\ldots,k$, and obtained a *standard sequence* of outcomes $x^0, x^1,\ldots, x^k$ such that successive elements of the sequence were equally spaced in terms of utility: $u(x^{j+1}) - u(x^j) = u(x^1) - u(x^0)$ for $j = 2,\ldots,k-1$. It is easily verified that if $G \succ g$ then the standard sequence is *increasing*, i.e. $x^j \succ x^{j-1}$ for $j = 1,\ldots,k$. If $G \prec g$ then the standard sequence is *decreasing*, i.e. $x^j \prec x^{j-1}$ for $j = 1,\ldots,k$.

Appendix 3A shows that if all temporal profiles involve the same unit of time and have common final periods then we can freely choose the utility of two outcomes. Since we only considered such temporal profiles, we set $u(x^k) = 1$ and $u(x^0) = 0$ for increasing standard sequences, yielding $u(x^j) = j/k$ for $j = 0,\ldots,k$. For decreasing standard sequences, we set $u(x^k) = -1$ and $u(x^0) = 0$, yielding $u(x^j) = -j/k$ for $j = 0,\ldots,k$.

Second stage: measurement of the time weights

Once the utility function is known, the measurement of the time weights is straightforward. We elicited the outcome z such that a participant was indifferent between $z_0 x^0_t$, i.e. z now and x^0 at time point t, and $x^0_0\, x^k_t$, i.e. x^0 now and x^k at time point t. By (1) we obtain that

$$u(z) + \lambda_t u(x^0) + \sum_{s\neq 0,t} \lambda_s u(0) = u(x^0) + \lambda_t u(x^k) + \sum_{s\neq 0,t} \lambda_s u(0) \qquad (5)$$

and, hence, by the adopted scaling, $|u(z)| = \lambda_t$. By varying t, we could elicit different time weights. The elicited outcomes z typically did not belong to the standard sequence elicited in the first stage and their utility was unknown. If participants have positive time preference, however, then z will lie between two elements of the elicited standard sequence and we could approximate the utility of z through the known utility of these elements of the standard sequence. This approximation will be good if successive elements of the standard sequence are not too far apart. We return to the issue of approximation below.

3.4 Experiment

The aim of the experiment was to elicit the intertemporal utility function and the time weights both for gains and for losses through the procedure outlined above.

Participants and incentives

Seventy participants were recruited and were paid a fixed amount of €12.50 to join the experiment. The participants were students from different faculties of the Erasmus University Rotterdam. Before the actual experiment, we tested the design in several pilot sessions using other students and university staff as participants.

Throughout the experiment we used hypothetical choices. There were several reasons for using hypothetical instead of real incentives. A first reason was the problem in organizing payments in the future, some of which occurred in four years time. Second, because utility tends to be close to linear for small amounts (Wakker and Deneffe, 1996), we used large money amounts to capture the effect of utility curvature. Paying these amounts for real would have been prohibitively expensive. Third, there were ethical constraints to use real incentives for the losses part of the experiment. Finally, in hypothetical questions one can ask participants to assume that there is no risk associated with future payments. With real stakes, participants may consider the receipt of future money amounts uncertain, which could inflate the discounting of these amounts.

Some studies have investigated the differences between real and hypothetical money amounts in intertemporal decision making, but no clear evidence exists that hypothetical amounts are discounted differently than real amounts (Frederick et al., 2002). In decision under uncertainty, real and hypothetical incentives do not seem to give qualitatively different results, although real incentives tend to reduce data variability (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001).

Procedure

The experiment was administered with each participant individually using a computer program. Answers were entered into the computer by the interviewer, so that participants could concentrate on the questions and mistakes could be reduced. Each individual session lasted between 30 and 45 minutes. Throughout the experiment, participants were encouraged to think aloud to obtain insight into the reasoning underlying their answers.

All indifferences were elicited through a sequence of choices. We used choices because empirical evidence suggests that choice-indifferences lead to fewer inconsistencies than indifferences determined by matching, where participants are directly asked to state their indifference value (Bostic et al., 1990). Because we used choices, our study employs what Shelley (1993) refers to as a neutral frame and, hence, we could test whether the gain-loss asymmetry was due to a framing effect.

The interviewer used a scroll bar to vary the value of the outcome that we sought to elicit, starting with values for which preferences were clear and then “zooming in” on the indifference value. Examples of the computer screens that participants faced in the first and the second stage of the experiment are in Appendix 3B and 3C.

We elicited the general discounted utility model first for gains and then for losses. We always started with the gains part because we learnt from the pilot sessions that this made it easier for participants to understand the choice task. Both parts were preceded by a practice question. Recall from Section 3.3 that our method involved the selection of two gauge outcomes denoted G and g. For gains (losses), G was a prospect giving a 50% chance of winning (losing) €2000 (and nothing otherwise) and g was a prospect giving a 50% chance of winning (losing) €500. Hence, we elicited increasing standard sequences for gains and decreasing standard sequences for losses. We used risky prospects instead of riskless money amounts to discourage heuristics like simply computing the difference in absolute values, which we observed in pilot studies where riskless money amounts were used. It is important to emphasize that our results are robust to participants' evaluation of prospects (e.g. according to expected utility or prospect theory) provided that the same theory is used throughout the experiment. As mentioned before, we chose substantial amounts of money to be able to detect utility curvature. A few participants mentioned budgetary constraints in the losses questions. They were told to assume that an interest-free loan was available to pay off the losses.

The starting outcome x^0 was €0 both for gains and for losses. The delay t was set equal to 1 year in the elicitation of utility. Hence, in the first question we elicited the money amount x^1 that made participants indifferent between prospect G at time point 0 and nothing in 1 year and prospect g at time point 0 and x^1 in 1 year. Both for gains and for losses we elicited 6 elements of the standard sequence.

In the second part of the experiment, we elicited the time weights for t = 3 months, 6 months, 1 year, 2 years, 3 years, and 4 years, both for gains and for losses. The order in which the time weights were elicited was random. Both for gains and for losses, we tested for consistency by repeating the first elicitation at the end of each experimental task. That is, in the elicitation of utility, we repeated the elicitation of x^1 after $x^1,\ldots,x^6$ had been elicited and in the elicitation of the time weights we repeated the elicitation of the time weight that had been elicited first[6] after the time weights for 3 months,…, 4 years had been elicited.

[6] Recall that the order in which the time weights were elicited was random.

Analyses

The results for means and medians were similar and, hence, we will only report the medians in the analysis of the aggregate data. Due to the presence of outliers, we focused on nonparametric tests to test for statistical significance.

To investigate the curvature of utility at the individual level, we computed

$$\partial_j = (x_{j+1} - x_j) - (x_j - x_{j-1}),\ j = 1,\ldots,5, \tag{6}$$

i.e. how much successive outcome intervals increase or decrease. We observed ten values of ∂_j for each participant, five for gains and five for losses. For gains, a positive value of ∂_j corresponds to a concave part of the utility function. A positive ∂_j means that an individual needs a larger increase in money to obtain a given increase in utility (1/6) at higher amounts than he needs at lower amounts. Likewise, a negative value of ∂_j corresponds to a convex part of the utility function and a value of zero to linear utility. For losses, a positive value of ∂_j corresponds to a convex part of the utility function and a negative value of ∂_j to a concave part.

Both for gains and for losses, we determined for all five ∂_j's of each participant whether it corresponded to a concave, convex or linear part of the utility function. We classified a participant as having linear (concave, convex) utility if he had at least three linear (concave, convex) parts. Again, we did this both for gains and for losses. We used a criterion of three instead of five similar parts, to account for response error.[7] If none of the three parts (linear, concave or convex) occurred more than twice, the participant was left unclassified.

To smoothen out irregularities in the data, we also analyzed the data under specific parametric assumptions about utility. We examined two parametric families: the power family and the exponential family. Both are widely used in economics and decision analysis. Because the two functions yielded similar results we will only report the results for the power family. Let $z = x/x^6$, $x \in [0,x^6]$. The *power family* is defined by $|z|^r$ if $r > 0$, by $\ln(z)$ if $r = 0$, and by $-|z|^r$ if $r < 0$. For gains (losses), $r < 1$ corresponds to concave (convex) utility and $r > 1$ to convex (concave) utility; the case $r = 1$ corresponds to linear utility both for gains and for losses. We estimated the parametric families both for the median data and for each individual separately. The estimation was by nonlinear least squares.

We used the estimates of the power coefficients to obtain another, parametric, classification of individual participants. For gains (losses) we classified a participant as concave (convex) if his power coefficient was below 0.95, as linear if his power coefficient was between 0.95 and 1.05, and as convex (concave) if his power coefficient exceeded 1.05.

To compute the time weights we had to approximate the utility of the elicited outcome z in (5). We did this by linear approximation. We also used approximation by the estimated power and exponential utility. This affected the results only marginally and we do not report these results separately.

---

[7] Similar criteria were used by Fennema and van Assen (1999), Abdellaoui (2000), Etchart-Vincent (2004), and Abdellaoui et al. (2005).

From the elicited time weights we could estimate implied annual discount rates ρ_s as follows:

$$\lambda_s = \frac{1}{(1+\rho_s)^s}, \tag{7}$$

where s is time in years. We could then test whether the implied annual discount rates were constant. We computed the difference between the implied annual discount rates for adjacent time periods[8] and, hence, obtained five observations for each participant both for gains and for losses. If at least three of these observations were positive then the participant was classified as decreasingly impatient, i.e. as having decreasing discount rates over time, if at least three observations were negative then he was classified as increasingly impatient, and if at least three observations were zero then he was classified as a constant discounter. Again, we used a criterion of three out of five to account for response error.

We also used the elicited time weights to estimate the parameter(s) in constant discounting, generalized hyperbolic discounting, proportional discounting, power discounting, and quasi-hyperbolic discounting. Each model was estimated separately, so we did not assume that ρ in constant discounting and ρ in quasi-hyperbolic discounting were equal, that γ in generalized hyperbolic discounting and γ in proportional discounting were equal or that α in generalized hyperbolic discounting and α in power discounting were equal. The models were estimated by nonlinear least squares both for the median data and for each participant separately. To test whether the results were sensitive to the specification of the unit of time, we performed the estimations for different specifications of the unit of time (years, months, and weeks).

Goodness of fit was assessed by Akaike's information criterion. An important advantage of this criterion is that it takes into account that the discounting models differ in the number of parameters employed. The fit of nested models was also compared through likelihood ratio tests. Because the coefficients for the median data were very close to the medians of the estimated coefficients for the individual data, we will focus on the individual data.

3.5 Results

The data of two participants were excluded from the analysis, because these participants gave answers that did not correspond to their reasoning. The data of another participant were lost due to a computer crash. As a result, the data of 67 individuals (31 females) were included in the analysis. The consistency of the data was good: none of the tests that we performed revealed significant inconsistencies in participants' responses ($p > 0.05$ in all tests). The individual parametric estimates are given in Appendix 3D.

---

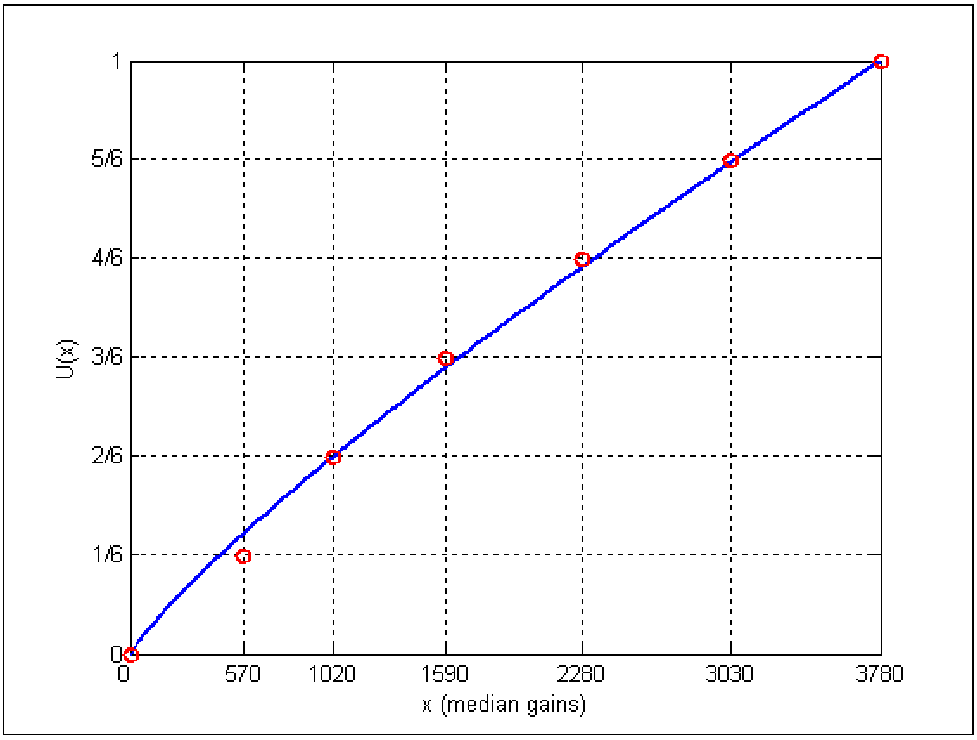
[8] That is, we computed $\rho_{3months} - \rho_{6months}$, $\rho_{6months} - \rho_{1year}$, $\rho_{1year} - \rho_{2years}$, $\rho_{2years} - \rho_{3years}$, and $\rho_{3years} - \rho_{4years}$.

3.5.1 Gains

Utility

Concavity of utility was most common at the individual level. Twenty-two participants were classified as concave, 16 as linear, and 7 as convex. The proportion of concave participants was significantly higher than the proportion of convex participants ($p < 0.01$). The classification based on the individual estimates of the power function also showed predominant concavity: thirty-eight individuals were classified as concave, 12 as linear and 17 as convex. The difference between the proportion of concave and the proportion of convex participants was significant ($p < 0.01$). The median of the individual estimates of the coefficient in the power utility function was 0.91 (interquartile range (IQR) = 0.76−1.05), which indicated slight concavity. This median was, however, not significantly different from 1 ($p = 0.075$).

FIGURE 3.1. THE UTILITY FUNCTION FOR GAINS BASED ON THE MEDIAN DATA.



Our findings on utility were comparable to those observed for decision under risk and uncertainty. The proportions of concave and convex participants were slightly lower and the proportion of linear subjects was slightly higher (Abdellaoui, 2000; Abdellaoui et al., 2005). The median power coefficient that we observed was similar to studies that estimated the utility for gains in decision under risk and uncertainty. Abdellaoui et al. (2005) and Schunk and Betsch (2006), for example, also found a median estimate of 0.91 for decision under uncertainty, whereas for decision under risk this estimate was 0.88 in Tversky and Kahneman (1992) and 0.89 in Abdellaoui (2000).

The aggregate data also showed concavity for gains. Figure 3.1 displays the utility function for gains based on the median data. The x-axis shows the medians of the elicited elements of the standard sequences for gains, the y-axis their utility. The difference between successive elements of the standard sequence generally increased, consistent with concave utility. The hypothesis that the difference between successive elements of the standard sequence was constant, the case corresponding to linear utility, could be rejected ($p < 0.01$). The estimated power coefficient for the median data was 0.84 (White's corrected standard error = 0.027), which differed significantly from 1 ($p < 0.01$), the case corresponding to linear utility. Figure 3.1 also shows the estimated power function. As the figure shows, the fit of the estimation was very good.

Time weights

The individual level data showed clear evidence of decreasing impatience: 55 participants were classified as decreasingly impatient and only 12 as increasingly impatient. The difference between the proportion of decreasingly impatient participants and the proportion of increasingly impatient participants was significant ($p < 0.01$).

Table 3.1 shows the median time weights for each delay, as well as the median annual discount rates that were implied by these weights. The median implied discount rates were low in comparison with the rates estimated in most previous studies. The median discount rates declined over time, which is consistent with decreasing impatience. The pattern of decreasing impatience was significant: the hypothesis that the implied annual discount rates were constant could be rejected ($p < 0.01$).

TABLE 3.1. TIME WEIGHTS AND IMPLIED ANNUAL DISCOUNT RATES FOR GAINS. INTERQUARTILE RANGES IN PARENTHESES.

Delay	3 months	6 months	1 year	2 years	3 years	4 years
Median time weight	0.970 (0.942-0.990)	0.944 (0.897-0.978)	0.923 (0.787-0.953)	0.831 (0.672-0.915)	0.735 (0.487-0.891)	0.735 (0.352-0.870)
Annual discount rate	12.8% (4.0%-26.9%)	12.3% (4.6%-24.3%)	8.3% (4.9%-27.1%)	9.7% (4.5%-22.0%)	10.8% (3.9%-27.1%)	8.0% (3.6%-29.9%)

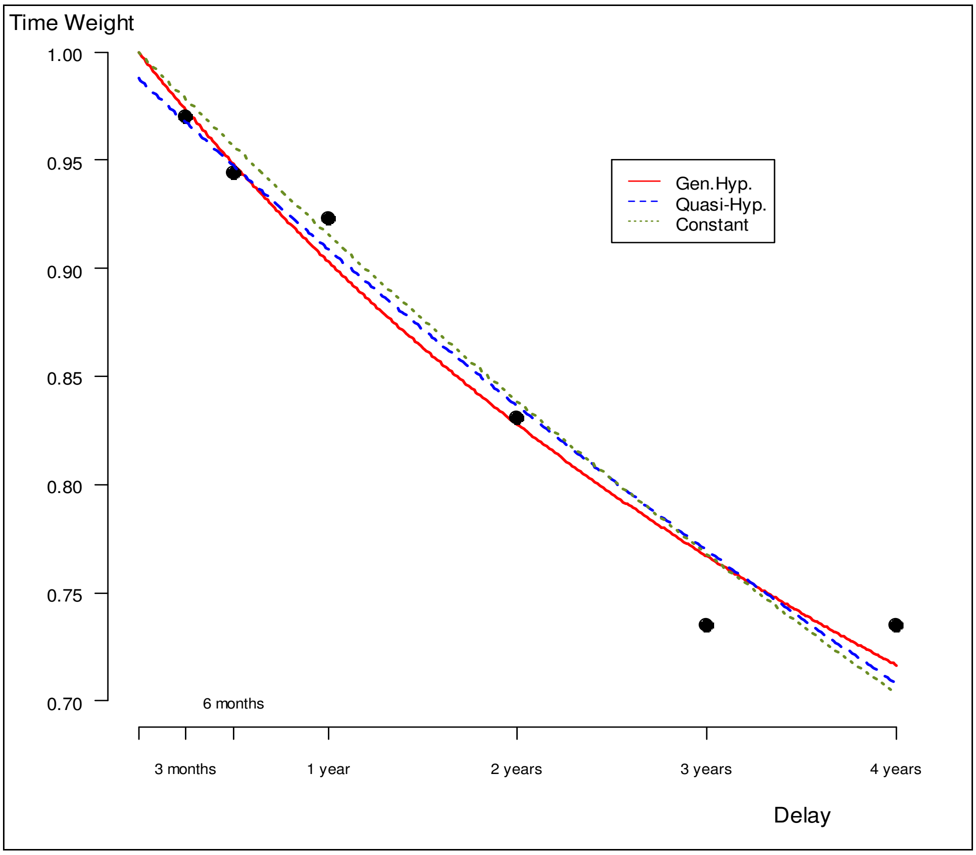
Table 3.2 shows the medians of the estimated individual parameters for each of the five discounting models. Generalized hyperbolic discounting fitted the data best. The difference in goodness of fit between generalized hyperbolic discounting and the other models was always significant. The fit of the other four models was similar and no significant differences were observed. The results were not sensitive to the specification of the unit of time.

TABLE 3.2. MEDIANS OF THE INDIVIDUAL PARAMETER ESTIMATES FOR THE DISCOUNTING MODELS (GAINS).

Model	Constant	Proportional	Power	Generalized Hyperbolic		Quasi-hyperbolic	
Parameter	δ^+	γ^+	α^+	γ^+	α^+	β^+	δ^+
Median (IQR)	0.102 (0.038 - 0.263)	0.117 (0.039-0.309)	0.213 (0.080-0.460)	0.291 (–0.087-1.952)	0.130 (0.070-0.316)	0.988 (0.958-1.009)	0.078 (0.033-0.208)

The median value of γ^+ in generalized hyperbolic discounting shows limited deviation from constant discounting. The individual estimates of γ^+ varied substantially, however. The wide variation of γ^+ was caused by poor convergence of the estimation algorithm for some participants. We could reject the hypotheses that γ^+ was equal to 0, the case of constant discounting, or equal to 1, the case of power discounting, ($p < 0.01$ in both cases) and also the hypothesis that $\alpha^+ = \gamma^+$, the case of proportional discounting ($p < 0.01$). The parameter β^+ in quasi-hyperbolic discounting was close to 1, but significantly lower than 1 ($p < 0.01$), suggesting a small but significant immediacy effect. There were only 17 subjects for whom β^+ was less than 0.95, which illustrates that for most subjects the immediacy effect was small. Both δ^+ in constant discounting and δ^+ in quasi-hyperbolic discounting ($p = 0.02$) and α^+ in power discounting and α^+ in generalized hyperbolic discounting ($p = 0.03$) differed significantly.

FIGURE 3.2. MEDIAN TIME WEIGHTS FOR GAINS AND FIT OF PARAMETRIC MODELS.



The estimation results based on the median data were largely similar. The main exception was that the immediacy effect was no longer significant. Figure 3.2 shows the fit of generalized hyperbolic discounting, quasi-hyperbolic discounting and constant discounting to the median data. The results of power discounting and proportional discounting were similar but are not displayed to keep the figure tractable. The figure shows that even though generalized hyperbolic discounting provided the best fit, the differences in fit between the models were limited.

3.5.2 Losses

Utility

The individual data showed no clear pattern in the direction of concave or convex utility. There were 122 concave parts in total, 98 linear parts, and 125 convex parts. Twenty-two participants had convex utility for losses and 20 had concave utility (9 had linear utility; the others could not be classified). The proportion of participants with convex utility did not differ significantly from the proportion with concave utility ($p > 0.10$). The classification based on the estimates of the power coefficient showed more evidence of convex utility for losses: 30 participants had convex utility, 23 linear utility, and 14 concave utility. The proportion of participants with convex utility was now significantly higher than the proportion with concave utility for losses ($p = 0.01$). The median of the individual estimates of the coefficient in the power utility function was 0.96 (interquartile range = 0.82-1.04), also indicating slight convexity for losses. This median was not significantly different from 1 ($p = 0.085$).

Our findings on utility for losses were also similar to those obtained for decision under risk and under uncertainty. The proportion of convex participants was similar, the proportion of concave participants was somewhat higher and the proportion of linear participants somewhat lower (Fennema and van Assen, 1999; Abdellaoui, 2000; Etchart-Vincent, 2004; Abdellaoui et al., 2005). Our median estimate for the power coefficient of 0.96 was close to the estimates of 0.88 in Tversky and Kahneman (1992), 0.84 in Fennema and van Assen (1999), 0.92 in Abdellaoui (2000), and 0.96 for small losses and 0.98 for large losses in Etchart-Vincent (2004).

FIGURE 3.3. THE UTILITY FUNCTION FOR LOSSES BASED ON THE MEDIAN DATA.

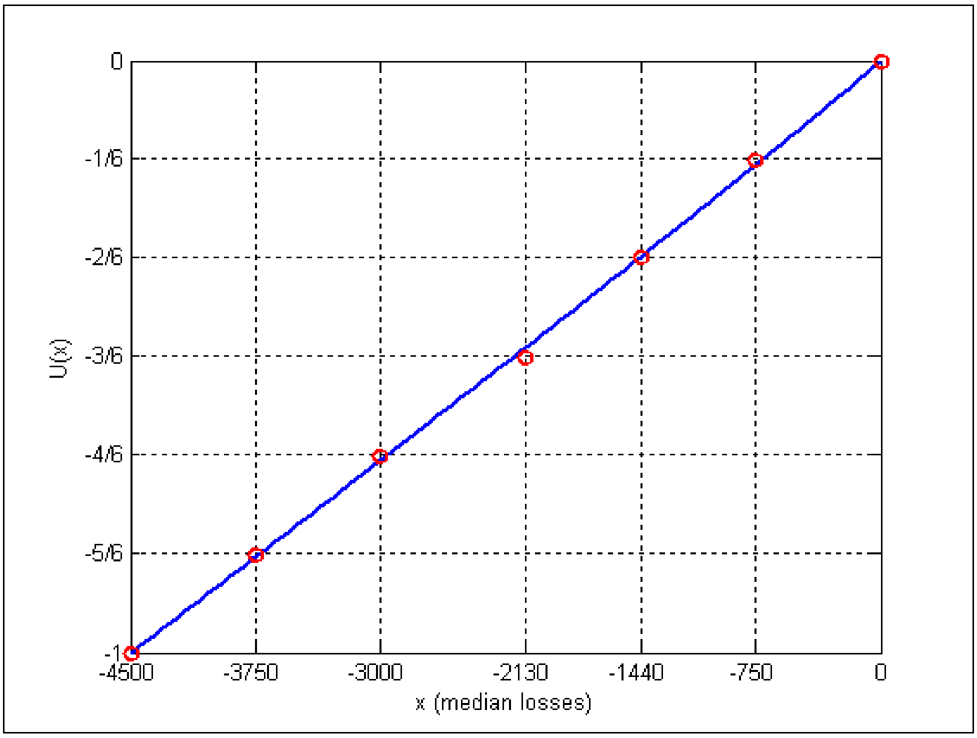


Figure 3.3 shows the utility for losses based on the median data. The x-axis shows the medians of the elements of the standard sequences. The differences between successive elements of the standard sequence were close and, hence, the utility for losses was close to linear at the aggregate level. We could reject, however, the

null hypothesis that the differences between successive elements of the standard sequence were all equal ($p < 0.01$) and, hence, the hypothesis of linear utility.

The estimate of the power coefficient was 0.97 (White's corrected standard error: 0.016), which indicated slight convexity and which was just significantly different from 1 ($p = 0.05$). Figure 3.3 also shows the plot of the estimated power function. The fit of the estimation was very good.

Time weights

As for gains, decreasing impatience was the most common pattern at the individual level: 47 participants were decreasingly impatient, 18 increasingly impatient, and 2 participants were constant discounters. The proportion of decreasingly impatient participants was significantly higher than the proportion of increasingly impatient participants ($p < 0.01$).

Table 3.3 shows the median time weights and the implied annual discount rates for each of the six delays. The discount rates declined over time, as predicted by decreasing impatience, but the decline was modest. The pattern of decreasing impatience was significant, however: the null hypothesis that the implied discount rates were all equal could be rejected ($p < 0.01$).

TABLE 3.3. TIME WEIGHTS AND IMPLIED ANNUAL DISCOUNT RATES FOR LOSSES. INTERQUARTILE RANGES IN PARENTHESES.

Delay	3 months	6 months	1 year	2 years	3 years	4 years
Median time weight	0.984 (0.967-0.991)	0.969 (0.928-0.989)	0.947 (0.893-0.968)	0.898 (0.817-0.941)	0.871 (0.752-0.932)	0.834 (0.652-0.909)
Annual discount rate	6.8% (3.8%-14.6%)	6.6% (2.3%-16.1%)	5.6% (3.3%-12.0%)	5.5% (3.1%-10.7%)	4.7% (2.4%-10.0%)	4.6% (2.4%-11.3%)

Table 3.4 shows the estimation results for the five discounting models at the individual level.[9] As for gains, generalized hyperbolic discounting provided the best fit. It fitted the data significantly better than the other models except that for power discounting the results are ambiguous. Based on Akaike's information criterion generalized hyperbolic discounting did not fit significantly better than power discounting ($p = 0.181$). However, based on the likelihood ratio test we could reject the hypothesis that γ^- was equal to 1 ($p < 0.01$).

Proportional discounting fitted significantly worse than all other theories ($p < 0.01$ in all cases). No significant differences in goodness of fit were observed between constant discounting, power discounting, and quasi-hyperbolic discounting. The conclusions about the relative performance of the different theories were not sensitive to the specification of the unit of time.

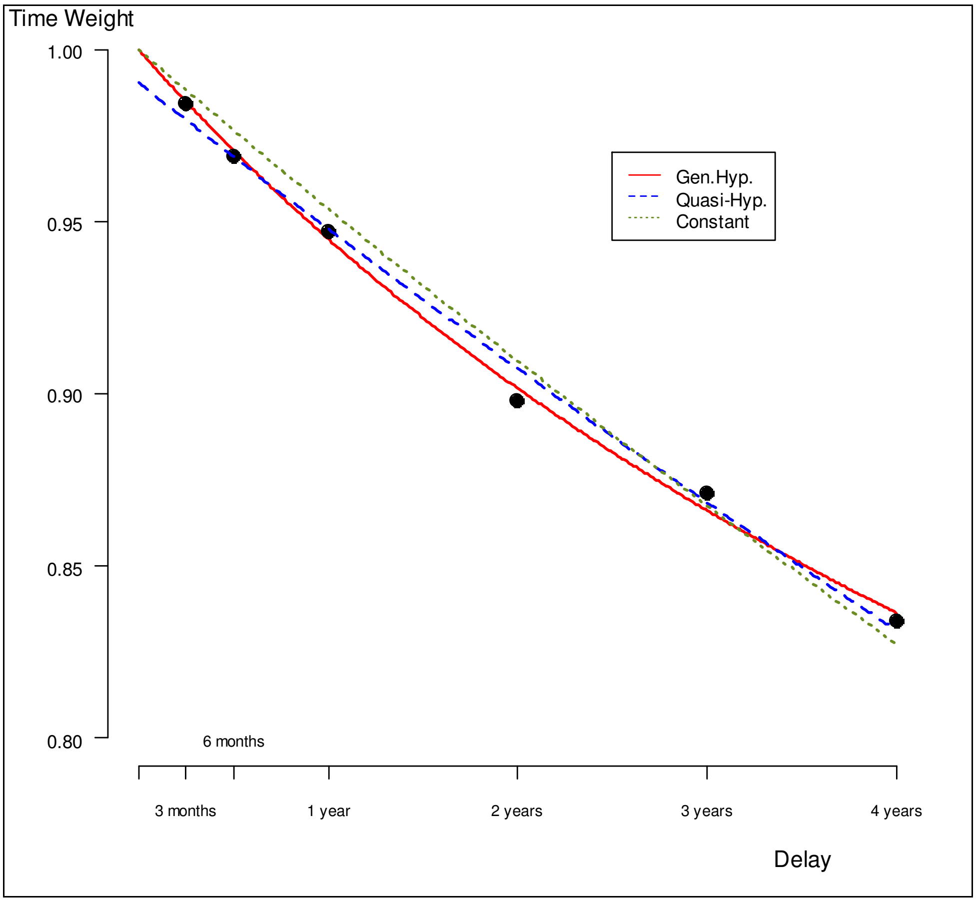
[9] For three participants, it was not possible to estimate the models because they did not discount losses.

TABLE 3.4. INDIVIDUAL PARAMETER ESTIMATES FOR THE DISCOUNTING MODELS (LOSSES).

Model	Constant	Proportional	Power	Generalized hyperbolic		Quasi-hyperbolic	
Parameter	ρ^-	γ^-	α^-	γ^-	α^-	β^-	ρ^-
Median (IQR)	0.056 (0.032-0.101)	0.090 (0.053-0.182)	0.120 (0.070 - 0.208)	0.512 (–0.074 -1.555)	0.100 (0.053-0.242)	0.989 (0.969-1.003)	0.046 (0.029-0.090)

The estimate for γ^- in generalized hyperbolic discounting indicated stronger deviations from constant discounting than we observed for gains. The estimates varied substantially across individuals, however. We could reject the hypotheses that γ^- was equal to 0, 1 or to α^- ($p < 0.01$ in all cases). The estimate for β^- suggested a small but significant immediacy effect ($p < 0.01$). Only 15 subjects had a value of β^- smaller than 0.95, indicating that for most subjects the immediacy effect was modest. The parameters ρ^- under constant discounting and ρ^- under quasi-hyperbolic discounting differed significantly ($p < 0.01$); the parameters α^- under power discounting and α^- under generalized hyperbolic discounting did not differ significantly ($p > 0.10$).

FIGURE 3.4. MEDIAN TIME WEIGHTS FOR LOSSES AND FIT OF PARAMETRIC MODELS.



The estimation results based on the median data were largely similar. The main difference was that the immediacy effect was no longer significant and the estimate of γ^{-} in generalized hyperbolic discounting indicated less deviation from constant discounting and was only 0.228. This estimate differed significantly from 0, however ($p = 0.027$). Figure 3.4 shows the fit to the median data of generalized hyperbolic discounting, quasi-hyperbolic discounting, and constant discounting. Generalized hyperbolic fitted the data best, but the figure shows that the fit of quasi-hyperbolic discounting and, to a lesser degree, constant discounting was also good. All parametric models fitted the data better for losses than for gains.

3.5.3 Comparison of gains and losses

In order to check whether there was a gain-loss asymmetry, we investigated the difference between the time weight for gains and the time weight for losses for each delay. As Tables 3.2 and 3.4 show, the time weights were higher for losses than for gains. The difference was significant for all delays ($p = 0.02$ for 3 months, $p < 0.01$ for 6 months, $p = 0.03$ for 1 year, $p < 0.01$ for 2 years, 3 years, and 4 years).

Mixed results obtained when we compared the parameters in the five discounting models for gains and for losses. Remember that none of the discounting models allows for a difference in these parameters. The estimates in constant discounting and power discounting differed significantly for gains and for losses ($p < 0.01$). The parameters in generalized hyperbolic discounting and proportional discounting did not differ significantly for gains and for losses ($p > 0.10$ in all tests). In the quasi-hyperbolic model, finally, the estimates of ρ^+ and ρ^- differed significantly ($p < 0.01$), whereas β^+ and β^- did not differ significantly ($p > 0.10$).

3.5.4 The effect on the time weights of assuming linear utility

As mentioned before, most previous studies that estimated discount rates assumed linear utility. To assess the bias resulting from assuming linear utility, we also computed the time weights under the assumption that utility was linear and compared these with the parameter-free time weights obtained without making assumptions about utility. The median time weights were lower under linear utility and, hence, the annual discount rates were higher. However, the time weights under linear utility did not differ significantly from the parameter-free time weights ($p > 0.05$) with the exception of the time weight for 2 years for gains ($p = 0.05$). The gain-loss asymmetry became more pronounced when linear utility was assumed ($p = 0.01$ for 3 months, $p = 0.02$ for 6 months, $p < 0.01$ for the other delays).

The fit of the discounting models was significantly better when we used the utility-adjusted time weights than when we used the time weights that were computed under the assumption that utility was linear. For gains, the exception was proportional discounting ($p = 0.056$), for losses the exceptions were power discounting ($p = 0.063$) and generalized hyperbolic discounting ($p = 0.563$). The conclusions about the relative fit of the discounting models were hardly affected by assuming linear utility. The only differences were that for losses generalized hyperbolic discounting now fitted significantly better than all other models and that proportional discounting no longer fitted significantly worse than constant discounting, quasi-hyperbolic discounting, and power discounting.

3.6 Discussion

Loewenstein and Prelec (1992), extending Kahneman and Tversky (1979) to intertemporal choice, suggested that utility be concave for gains and convex for losses. We found some evidence for their proposition. Our data were consistent with concave utility for gains, but for losses the picture was less clear, although the predominant shape of utility was slightly convex. The power coefficients that we estimated were consistent with Loewenstein and Prelec's (1992) assumption that utility is more elastic for losses than for gains.

Interestingly, our findings on the degree of utility curvature were close to those obtained in decision under uncertainty and in decision under risk. This held both for gains and for losses. While requiring further evidence, this finding may suggest that there exists one unifying concept of utility. Economists have traditionally argued that utility differs across domains and, hence, that the utility function that is relevant for decision under risk cannot be employed in other contexts, such as intertemporal decision making. In applied economics transferability of utility is, however, commonly assumed. For example, in health economics measurements of utility under risk are routinely used in welfare comparisons. Our findings provide some tentative support for the transferability of utility that is commonly assumed in applied economics.

The discount rates we observed were lower than the rates observed in most previous studies (Frederick et al., 2002). This could not be entirely explained by the fact that, contrary to previous studies, we made no assumptions about utility, because even when linear utility was imposed the observed discount rates were still relatively low. One explanation may be that our experiment was choice-based, whereas most previous studies used matching tasks. It is well-known that choice tasks tend to produce different results than matching tasks (Tversky et al., 1988). Ahlbrecht and Weber (1997) also observed different discounting patterns in choice than in matching. That said, even though our experimental tasks involved choices, it was clear to participants that we were looking for indifferences and in that sense our task resembled matching (Fischer et al., 1999). Empirical evidence shows that eliciting indifferences through a series of choices, as we did, produces results in between choice and matching (Delquié, 1997).

Of the main discounting models that we considered, generalized hyperbolic discounting fitted the data best. The fit of constant discounting was rather good and we could not conclude that quasi-hyperbolic discounting, power discounting or proportional discounting fitted the data unambiguously better than constant discounting. One reason why quasi-hyperbolic discounting is used so much in applications is the alleged belief that it fits individual choice behavior better than constant discounting. Our data provide little support for this belief. If the aim of models is to accurately describe individual intertemporal choice behavior then the most appropriate model to use seems generalized hyperbolic discounting. Estimation of the generalized hyperbolic model is no more complicated than quasi-hyperbolic discounting and the fit was significantly better. It should be emphasized, however, that there exist other reasons to use quasi-hyperbolic discounting besides descriptive accuracy, one of which is the greater tractability of the model in theoretical analyses. A cause of concern in our study is the instability of the estimates for the parameter γ in generalized hyperbolic discounting.

Our experiment used a neutral frame as did Shelley (1993) and Ahlbrecht and Weber (1997). Contrary to those studies, we observed evidence for a gain-loss asymmetry and, hence, our data do not corroborate the conclusion that the gain-loss asymmetry is caused by a framing effect. The difference between our findings and those of Shelley (1993) and Ahlbrecht and Weber (1997) cannot be explained by the assumption of linear utility made in these studies, because assuming linear utility actually increased the gain-loss asymmetry.

The finding that assuming linear utility seemed not to cause serious biases is important for empirical research into intertemporal preferences. As mentioned before, most studies have hitherto assumed linear utility but it was not known to what extent this assumption distorted their findings. Our results suggest that this distortion was modest. This finding was not caused by our method for measuring utility. An easy heuristic to adopt in responding to the utility elicitation questions might be by keeping the difference between successive elements of the standard sequence constant, which would lead to linear utility. However, both for losses and for gains there were only six participants for whom the difference between successive elements of the standard sequence was always constant; there were only three participants for whom the difference between successive elements of the standard sequence was constant both for gains and for losses. These limited numbers suggest that the heuristic may not have caused serious biases.

Our method used chained measurements, i.e. answers from previous questions were used as inputs in later stages. A possible danger of using chained measurements is error propagation: errors in earlier responses get transferred to later responses. Bleichrodt and Pinto (2000) and Abdellaoui et al. (2005) examined the effect of error propagation in their studies and concluded that it had little impact. Since our method was based on a similar chaining process as theirs, their findings suggest that the effect of error propagation was limited in our study as well.

A crucial assumption in our method was that participants behaved according to the general discounting model (1). This model underlies all of the main discounting models used in the literature. A central property of (1) is intertemporal additivity. There is some evidence of violations of intertemporal additivity (Loewenstein and Sicherman, 1991; Frank and Hutchens, 1993). It is not clear how important these violations of (1) are. As mentioned by Loewenstein and Prelec (1992), they seem particularly relevant when evaluating complete alternative sequences of outcomes like savings plans or multiyear salary contracts. In our experiment, we considered, however, elementary types of intertemporal choices. We tried to mitigate the possible effect of violations of intertemporal additivity by using prospects in the elicitation of utility. We learned from pilot tests that using prospects made it more likely that people viewed things that happened at different points as separate and, hence, behaved more in line with intertemporal additivity.

The order of the tasks was fixed throughout the experiment. We always started with the gains part, because, as we observed in the pilot sessions, participants found this easier. It may be, although we do not consider this likely, that participants became more aware of their true preferences during the experiment and that this has caused the gain-loss asymmetry in our study. By the construction of our elicitation method, we always had to elicit utility prior to the elicitation of the time weights. We cannot think of any systematic bias that may have arisen because of this.

3.7 Conclusion

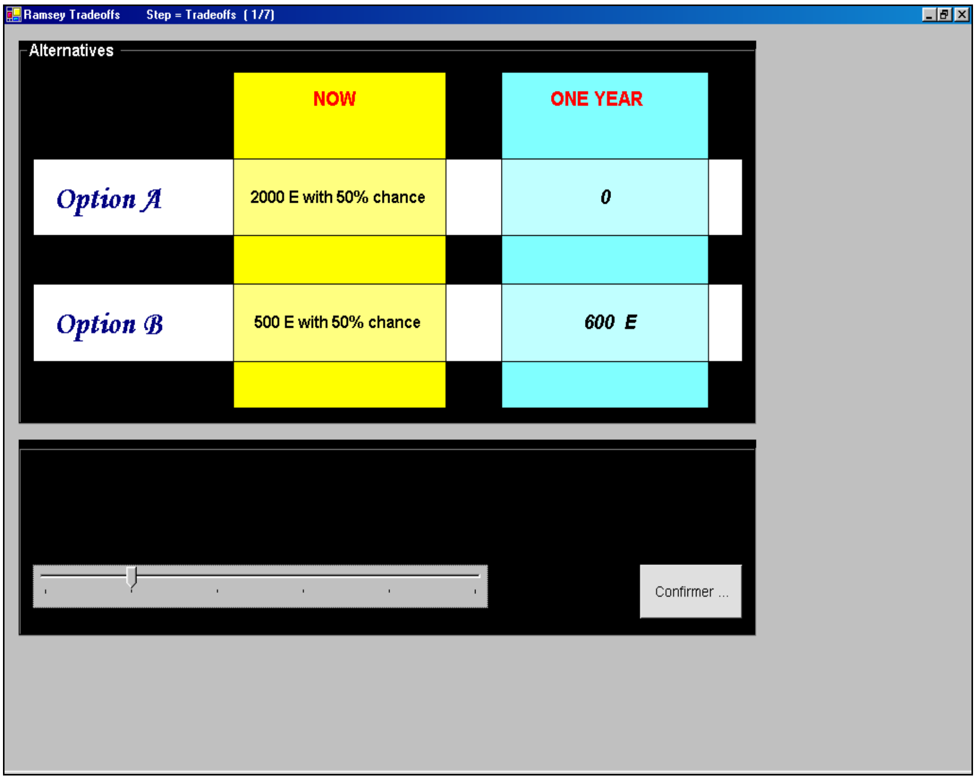
This chapter has presented a parameter-free method to measure the discounted utility model in its entirety. Hence, we are the first to measure the utility function in intertemporal choice and we provide more robust evidence on the discounting of monetary outcomes. We found concave utility for gains and slightly convex utility for losses, which supports a hypothesis put forward by Loewenstein and Prelec (1992). Utility in intertemporal choice was close to previously found results on utility in decision under risk and uncertainty suggesting the existence of one unifying concept of utility. Our data confirmed decreasing impatience. The decrease was, however, modest and the fit of constant discounting was rather good. Of the hyperbolic discounting models that we examined, generalized hyperbolic discounting fitted the data best. Our data were less supportive of the widely-used quasi hyperbolic discounting model: it did not fit significantly better than constant discounting and the immediacy effect that we observed was small. We found some evidence for a gain-loss asymmetry in the time weights, which contradicts earlier conclusions that the gain-loss asymmetry was due to a framing effect (Shelley, 1993; Ahlbrecht and Weber, 1997) and also contradicts Loewenstein and Prelec (1992) who suggested that the gain-loss asymmetry was a consequence of the shape of the utility function only. Finally, the assumption of linear utility seemed not to bias the estimated time weights and discount rates in a significant manner. Hence, our study suggests that assuming linear utility in future empirical studies and in practical applications may not be very harmful at least for qualitative purposes.

Appendix 3A. Proof that we can freely choose the utility of two outcomes when the unit of time and the final period are the same across temporal profiles

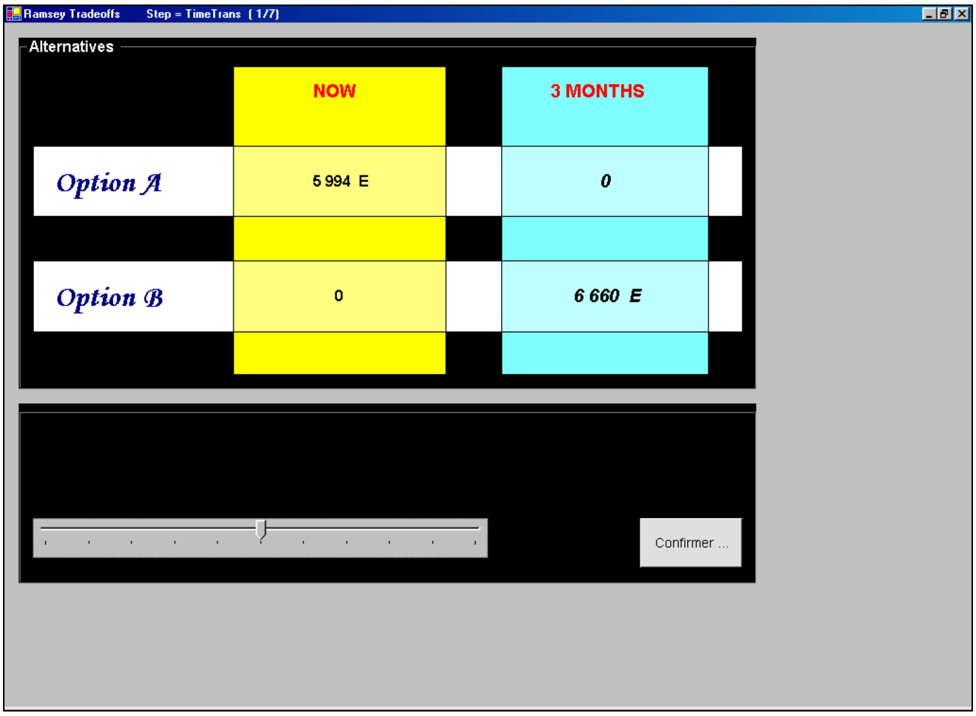
Because the unit of time and the final period are the same across temporal profiles, we restrict comparison to profiles x and y where the difference in timing between x_{t-1} and x_t is equal to the difference in timing between y_{t-1} and y_t and also the final periods are the same. Suppose that $\sum_{t=1}^{T}\lambda_t\, u(x_t) \geq \sum_{t=1}^{T}\lambda_t\, u(y_t)$ and that $\sum_{t=1}^{T}\lambda_t\, u(.)$ represents $\succcurlyeq$ over temporal profiles. If we replace u by $v = \alpha u + \beta$ with $\alpha > 0$ and β real, we obtain $\sum_{t=1}^{T}\lambda_t\, v(x_t) = \alpha\sum_{t=1}^{T}\lambda_t\, u(x_t) + \beta\sum_{t=1}^{T}\lambda_t \geq \alpha\sum_{t=1}^{T}\lambda_t\, u(y_t) + \beta\sum_{t=1}^{T}\lambda_t = \sum_{t=1}^{T}\lambda_t\, v(y_t)$. Hence, v also represents $\succcurlyeq$. Note that the assumption that the unit of time and the final period are the same across the profiles was crucial in the proof.

□

Appendix 3B. Example of the display participants faced in the elicitation of utility



Appendix 3C. Example of the display participants faced in the elicitation of the time weights



Appendix 3D. Individual parameter estimates

TABLE 3A.1. INDIVIDUAL PARAMETER ESTIMATES FOR GAINS.

Participant	ρ^+ (CD)	γ^+ (PD)	α^+ (PowD)	γ^+ (GHD)	α^+ (GHD)	ρ^+ (QHD)	β^+ (QHD)	r^+ (PowU)
1	0.055	0.058	0.111	-0.101	0.044	0.060	1.012	0.838
2	0.062	0.067	0.129	0.298	0.086	0.063	1.002	0.626
3	0.036	0.038	0.081	2.554	0.130	0.025	0.971	0.914
4	0.051	0.054	0.106	-0.065	0.044	0.047	0.989	0.920
5	0.013	0.013	0.028	-0.106	0.011	0.013	0.999	1.000
6	0.003	0.003	0.007	25.629	0.057	0.001	0.995	0.488
7	0.092	0.101	0.186	0.387	0.132	0.077	0.966	0.826
8	0.084	0.092	0.172	0.427	0.126	0.076	0.982	0.669
9	0.159	0.193	0.319	5.793	0.818	0.091	0.864	0.967
10	0.667	0.816	0.889	0.440	0.698	0.789	1.077	3.299
11	0.021	0.022	0.046	2.464	0.073	0.015	0.984	0.704
12	0.147	0.167	0.282	-0.083	0.118	0.138	0.983	1.267
13	0.196	0.227	0.357	-0.091	0.153	0.216	1.039	1.041
14	0.037	0.038	0.078	1.279	0.087	0.028	0.977	0.779
15	0.276	0.329	0.477	-0.131	0.193	0.322	1.076	1.143
16	0.308	0.407	0.547	3.184	0.960	0.187	0.817	2.388
17	0.010	0.011	0.023	3.627	0.046	0.007	0.991	0.799
18	0.037	0.039	0.079	0.207	0.048	0.035	0.995	0.948

Participant	ρ^+ (CD)	γ^+ (PD)	α^+ (PowD)	γ^+ (GHD)	α^+ (GHD)	ρ^+ (QHD)	β^+ (QHD)	r^+ (PowU)
19	0.008	0.008	0.017	-0.248	0.002	0.008	0.998	0.939
20	0.033	0.034	0.072	4.306	0.158	0.020	0.967	1.000
21	0.079	0.085	0.159	-0.070	0.067	0.078	0.997	0.688
22	0.025	0.025	0.054	1.479	0.064	0.019	0.986	0.808
23	0.174	0.203	0.329	0.017	0.165	0.173	0.998	1.000
24	0.443	0.584	0.704	0.662	0.607	0.323	0.858	0.943
25	0.138	0.159	0.272	0.450	0.203	0.110	0.941	0.759
26	0.038	0.040	0.079	0.106	0.043	0.036	0.995	0.689
27	0.040	0.042	0.085	1.541	0.104	0.031	0.977	1.592
28	0.726	1.148	1.003	61.840	12.300	0.197	0.577	1.011
29	0.223	0.267	0.410	-0.234	0.099	0.205	0.967	1.507
30	0.433	0.555	0.692	0.007	0.363	0.357	0.905	0.976
31	0.110	0.123	0.218	0.107	0.120	0.105	0.989	0.731
32	0.083	0.091	0.167	0.114	0.094	0.091	1.018	1.067
33	1.139	1.304	1.249	-0.024	0.746	1.296	1.064	2.525
34	0.027	0.027	0.058	2.959	0.100	0.018	0.978	0.883
35	0.250	0.292	0.437	-0.181	0.153	0.291	1.072	1.000
36	1295.3	20.525	7.697	50.388	36.460	2.077	0.201	3.277
37	0.324	0.368	0.530	-0.233	0.164	0.434	1.180	1.112
38	1.472	1.503	1.324	1.074	1.354	1.225	0.939	5.905
39	0.203	0.230	0.368	-0.247	0.075	0.197	0.988	0.820
40	0.376	0.440	0.595	-0.114	0.270	0.515	1.181	0.885
41	0.028	0.029	0.066	95160.1	602.89	0.001	0.931	0.786

Participant	ρ^+ (CD)	γ^+ (PD)	α^+ (PowD)	γ^+ (GHD)	α^+ (GHD)	ρ^+ (QHD)	β^+ (QHD)	r^+ (PowU)
42	0.112	0.129	0.229	1.243	0.252	0.088	0.946	1.075
43	0.578	0.761	0.843	0.583	0.703	0.458	0.890	0.730
44	0.151	0.174	0.289	0.057	0.152	0.162	1.022	1.451
45	0.102	0.117	0.213	3.515	0.407	0.066	0.918	1.018
46	0.125	0.138	0.238	-0.178	0.079	0.143	1.040	0.772
47	0.010	0.010	0.022	0.119	0.012	0.009	0.996	0.755
48	0.029	0.029	0.059	-0.031	0.027	0.032	1.008	0.694
49	0.040	0.042	0.085	0.561	0.068	0.034	0.984	0.824
50	0.167	0.194	0.315	0.044	0.165	0.172	1.010	0.780
51	0.357	0.418	0.578	-0.215	0.194	0.456	1.144	1.054
52	0.027	0.028	0.059	3.778	0.118	0.018	0.975	0.668
53	0.193	0.252	0.388	17.477	2.097	0.074	0.784	1.045
54	0.586	0.729	0.839	0.038	0.481	0.604	1.016	2.065
55	0.033	0.034	0.067	-0.140	0.024	0.035	1.005	1.000
56	0.049	0.052	0.104	0.551	0.083	0.040	0.976	0.594
57	0.041	0.042	0.081	-0.222	0.020	0.048	1.019	0.755
58	0.054	0.057	0.112	0.070	0.058	0.051	0.993	1.000
59	0.069	0.076	0.151	27.768	1.200	0.027	0.900	0.641
60	0.404	0.554	0.672	1.610	0.824	0.285	0.852	1.151
61	0.327	0.377	0.541	-0.247	0.141	0.413	1.141	0.729
62	0.165	0.192	0.315	-0.145	0.114	0.157	0.982	0.855
63	0.194	0.226	0.358	-0.098	0.149	0.205	1.021	2.969
64	0.076	0.083	0.156	0.292	0.103	0.070	0.986	0.732
65	0.072	0.079	0.154	2.294	0.231	0.050	0.948	0.774

Participant	ρ^+ (CD)	γ^+ (PD)	α^+ (PowD)	γ^+ (GHD)	α^+ (GHD)	ρ^+ (QHD)	β^+ (QHD)	r^+ (PowU)
66	0.030	0.031	0.064	0.548	0.051	0.026	0.991	0.845
67	0.331	0.447	0.576	2.375	0.859	0.211	0.833	0.736

Note: CD = constant discounting, PowD = power discounting, GHD = generalized hyperbolic discounting, QHD = quasi-hyperbolic discounting, PD = proportional discounting, PowU = power utility

TABLE 3A.2. INDIVIDUAL PARAMETER ESTIMATES FOR LOSSES.

Participant	ρ^- (CD)	γ^- (PD)	α^- (PowD)	γ^- (GHD)	α^- (GHD)	ρ^- (QHD)	β^- (QHD)	r^- (PowU)
1	0.072	0.115	0.153	2.102	0.219	0.052	0.950	1.027
2	0.026	0.038	0.059	3.331	0.110	0.018	0.976	0.732
3	0.018	0.027	0.038	-0.042	0.017	0.018	1.001	0.896
4	0.017	0.025	0.035	-0.230	0.008	0.017	0.999	0.799
5	0.017	0.025	0.037	0.730	0.033	0.015	0.993	0.911
6								0.792
7	0.084	0.138	0.173	0.367	0.121	0.074	0.974	0.971
8	0.009	0.014	0.021	2.329	0.032	0.007	0.993	0.810
9	0.045	0.069	0.094	0.263	0.060	0.040	0.989	0.634
10	0.061	0.093	0.135	3.694	0.267	0.037	0.943	1.157
11	0.014	0.020	0.033	25.542	0.247	0.006	0.978	0.897

Participant	ρ^- (CD)	γ^- (PD)	α^- (PowD)	γ^- (GHD)	α^- (GHD)	ρ^- (QHD)	β^- (QHD)	r^- (PowU)
12	0.100	0.168	0.197	-0.191	0.059	0.101	1.002	1.088
13	0.255	0.533	0.465	1.034	0.472	0.209	0.925	0.734
14	0.032	0.048	0.070	2.200	0.103	0.024	0.978	0.782
15								0.914
16	0.541	1.158	0.803	1.547	0.960	0.332	0.799	5.151
17	0.134	0.228	0.262	28.930	2.063	0.086	0.892	1.200
18	0.046	0.071	0.095	0.115	0.053	0.047	1.003	1.010
19	0.069	0.109	0.143	0.512	0.111	0.065	0.990	0.977
20	0.132	0.234	0.257	-0.002	0.124	0.128	0.990	1.415
21	0.089	0.145	0.184	1.208	0.200	0.067	0.949	1.023
22	0.037	0.056	0.079	0.751	0.070	0.031	0.985	0.805
23	0.132	0.231	0.252	-0.179	0.082	0.145	1.029	1.000
24	0.341	0.785	0.589	1.080	0.608	0.269	0.900	0.901
25	829.49	20.916	7.303	1162.94	382.88	0.330	0.150	1.759
26	0.037	0.057	0.076	-0.066	0.032	0.040	1.007	1.008
27	0.042	0.064	0.086	-0.046	0.038	0.042	1.002	1.285
28	0.371	0.794	0.610	-0.104	0.267	0.377	1.009	1.252
29	0.055	0.084	0.120	8.358	0.401	0.030	0.939	1.027
30	0.424	0.974	0.689	0.958	0.678	0.321	0.876	1.000
31	0.018	0.026	0.040	1.051	0.040	0.015	0.992	0.964
32	1.100	2.339	1.209	1.004	1.211	0.690	0.810	2.393
33	0.072	0.117	0.145	-0.074	0.061	0.075	1.006	1.355
34	0.085	0.139	0.173	0.306	0.116	0.085	0.999	1.047

Participant	ρ^{-} (CD)	γ^{-} (PD)	α^{-} (PowD)	γ^{-} (GHD)	α^{-} (GHD)	ρ^{-} (QHD)	β^{-} (QHD)	r^{-} (PowU)
35	0.061	0.098	0.126	0.070	0.066	0.064	1.005	1.000
36	0.015	0.020	0.034	6.819	0.100	0.008	0.981	0.735
37	0.068	0.106	0.139	0.256	0.089	0.070	1.005	0.586
38	0.037	0.056	0.076	-0.158	0.025	0.038	1.003	0.543
39	0.049	0.075	0.108	2.531	0.171	0.035	0.963	1.045
40	0.165	0.299	0.297	-0.233	0.078	0.223	1.123	0.653
41	0.013	0.019	0.028	-0.216	0.007	0.008	0.987	1.160
42	0.036	0.056	0.073	-0.174	0.023	0.038	1.006	0.930
43	0.038	0.059	0.077	-0.209	0.021	0.043	1.013	1.075
44								0.987
45	0.098	0.162	0.208	6.844	0.601	0.053	0.899	1.007
46	0.041	0.061	0.087	0.690	0.075	0.034	0.984	0.765
47	0.014	0.021	0.031	0.988	0.031	0.012	0.993	0.715
48	0.022	0.031	0.047	1.526	0.057	0.016	0.984	0.901
49	0.056	0.089	0.116	0.146	0.066	0.055	0.996	0.847
50	0.168	0.311	0.317	0.063	0.169	0.173	1.010	1.047
51	0.358	0.753	0.586	-0.132	0.246	0.411	1.074	0.952
52	0.052	0.080	0.112	1.317	0.126	0.041	0.971	0.893
53	0.101	0.170	0.208	1.483	0.249	0.077	0.945	0.780
54	0.203	0.380	0.373	-0.235	0.089	0.204	1.002	1.000
55	0.048	0.074	0.099	-0.022	0.045	0.044	0.991	1.000
56	0.073	0.120	0.146	-0.241	0.029	0.077	1.008	0.962
57	0.022	0.032	0.046	-0.102	0.017	0.018	0.990	0.961
58	0.059	0.093	0.120	-0.048	0.052	0.060	1.003	1.000

Participant	ρ^- (CD)	γ^- (PD)	α^- (PowD)	γ^- (GHD)	α^- (GHD)	ρ^- (QHD)	β^- (QHD)	r^- (PowU)
59	0.211	0.403	0.391	-0.099	0.160	0.180	0.942	1.276
60	0.083	0.134	0.168	0.142	0.096	0.073	0.978	1.633
61	0.272	0.521	0.459	-0.245	0.118	0.364	1.165	0.669
62	0.059	0.091	0.128	3.398	0.242	0.039	0.950	0.876
63	0.027	0.039	0.060	4.674	0.139	0.016	0.972	0.654
64	0.041	0.063	0.087	0.599	0.071	0.034	0.983	0.937
65	0.014	0.020	0.032	54.506	0.431	0.006	0.977	0.822
66	0.022	0.033	0.047	-0.090	0.019	0.022	1.000	0.986
67	0.043	0.064	0.092	1.555	0.113	0.033	0.974	0.938

Note: CD = constant discounting, PowD = power discounting, GHD = generalized hyperbolic discounting, QHD = quasi-hyperbolic discounting, PD = proportional discounting, PowU = power utility

4 Time standard sequences for quantifying and visualizing the degree of time inconsistency, using only pencil and paper[1]

Summary

This chapter introduces time standard sequences as a new tool for analyzing intertemporal preferences. Time standard sequences yield a new way to measure temporal discounting, while minimizing distortions due to violations of intertemporal separability. They make it particularly easy to observe and exactly quantify deviations from stationarity and the implied proneness to choice anomalies. Time standard sequences can easily be administered and analyzed, using only pencil and paper, and do not need any assumption about utility, or estimation thereof. They allow for the empirical discrimination between several hyperbolic discounting models that have been proposed in the literature as alternatives to constant discounting, such as quasi-hyperbolic, proportional, and generalized hyperbolic discounting. We tested the feasibility of time standard

[1] This chapter is based on Attema, Bleichrodt, Rohde, and Wakker (2006).

sequences in an experiment. Our findings suggest some new directions for theories of intertemporal choice.

4.1 Introduction

Many decisions involve tradeoffs over time. The most popular model for evaluating streams of outcomes over time, also assumed in this paper, is (general) discounted utility. In this model, streams of outcomes are evaluated by summing the discounted utilities of the outcomes received at various time points. In traditional approaches, the measurement of the discount function is difficult because both this function and the utility function are unknown parameters that have to be measured simultaneously (Coller et al., 2005). The difficulty is aggravated because separability of preferences over disjoint time periods, an assumption of the model, is extensively violated empirically, distorting assessments of discounting. We introduce time standard sequences as a new and simple tool for measuring discount functions. It turns out that we need no measurement of utility, or assumption about utility, to obtain the discount function. Distortions due to violations of time separability are minimized.

Samuelson (1937) introduced constant discounting, where a preference between two outcome streams does not change if all outcomes are delayed by an equal time period, reflecting constant impatience, a property also known as stationarity. Constant discounting has long been the standard for intertemporal choice in economics. An attractive feature is that, under some extra assumptions, constant discounting implies dynamically consistent behavior: plans for future decisions will be adhered to, and no arbitrage is possible.

Empirical evidence has revealed many violations of stationarity. Mostly, impatience is decreasing rather than constant (Frederick et al., 2002). People, who at present are not willing to wait for an improved but delayed outcome due to impatience, become willing to wait if all outcomes are delayed by the same amount of time. These people, thus, become less impatient as time proceeds. Under common assumptions, decreasing impatience implies dynamic inconsistency, which is usually considered irrational. All kinds of choice anomalies result, such as proneness to arbitrage.

Hyperbolic discounting models have been developed so as to model decreasing impatience. For example, quasi-hyperbolic discounting (Phelps and Pollak, 1968) assumes constant impatience for all future time points, but decreasing impatience at present. Then time inconsistency arises only if immediate consumption is involved. Generalized hyperbolic discounting (Loewenstein and Prelec, 1992) allows decreasing impatience at all time points. Analyses of traditional economic models change because of these new ways of discounting, and many previously unexplained phenomena can now be accommodated (Laibson, 1997). Hence, hyperbolic discounting is popular today.

Prelec (2004) introduced a theoretical measure of decreasing impatience, being the convexity index $-\frac{ln(\varphi)''}{ln(\varphi)'}$ of the logarithm of the discount function φ.[1] He demonstrated that this measure identifies different degrees of proneness to

[1] The index measures concavity for increasing functions, and convexity for decreasing functions such as φ.

inconsistencies and arbitrage when impatience is decreasing. He wrote: "Decreasing impatience provides a natural criterion for assessing whether a set of time preferences represents a more or less severe departure from the stationarity axiom. The criterion is associated with a simple normative diagnostic—the selection of inefficient (dominated) outcomes in two-stage decision problems" (p. 526).

At this stage, it may seem to be difficult to elicit or implement Prelec's measure in practice. It, apparently, first requires the measurement of the discount function and, next, taking the logarithm and determining its second derivative over its first. To measure the discount function, we, apparently, have to measure utility, or at least make assumptions about utility, because the discount function determines behavior only in combination with utility. Some analyses in the literature parametrically estimated utility and, subsequently, used these estimates to measure the discount function (e.g. Chapman, 1996a). Most analyses simply equated outcomes with utility, which amounts to the assumption of linear utility. Such assumptions can confound findings about discounting.

Time standard sequences provide a new way of directly measuring the degree of deviation from stationarity and the degree of time-inconsistent behavior. Surprisingly, we can immediately estimate Prelec's index $-\frac{ln(\varphi)''}{ln(\varphi)'}$ of time inconsistency and graphically depict it, using only pencil and paper, without need to carry out the measurements and calculations mentioned above. In particular, we need not determine the utility function. Through time standard sequences we can immediately tell who of two persons satisfies more decreasing impatience, and we can identify groups of people who are especially prone to losses and arbitrage because of time inconsistency, as we show in a representation theorem. Time standard sequences are easy to comprehend for participants, leading to reliable data.

Time standard sequences, together with one simple choice between outcome streams with two nonzero outcomes, completely identify the time discount function. Again, for this measurement no assumption about utility is needed. Our method is therefore, obviously, robust against distortions and nonlinearities in utility, and can be applied to general outcome sets, such as finite sets of qualitative health states.

We show how time standard sequences can test which of several hyperbolic models considered in the literature can be applied, and of those that can be, which best fit the choices of individuals. Until now, most studies only rejected constant discounting, but did not test which alternative was better. Exceptions are Abdellaoui et al. (2006), Keller and Strazzera (2002), and van der Pol and Cairns (2002). In an experiment, we demonstrate the feasibility of our method by measuring time standard sequences of 55 subjects.

Our experimental findings lead to a number of suggestions for new models of intertemporal choice. Several recent studies, discussed in Section 4.9, have found increasing impatience, which casts doubt on the universal decreasing impatience commonly assumed in time preference theories. Our study also finds a majority of increasing, rather than decreasing, impatience for the present and near future. After the present and near future, impatience becomes constant.

Most analyses of intertemporal discounting considered in the literature so far have focused entirely on decreasing impatience. The data of our study and some other recent studies suggest that the development of new tools for analyzing increasing impatience will be worthwhile. This point can be compared to the risk field, where tools for analyzing risk seeking are needed also if the majority of participants exhibit risk aversion. Without such flexibility of modeling, data fitting is not possible at the individual level.

We also find some fundamental violations of the general discounted utility model. This suggests that generalizations, primarily relaxing temporal separability, are desirable.

The outline of this chapter is as follows. Section 4.2 describes discounted utility and the various families of discount functions considered in this chapter. Time standard sequences and curves are presented in Section 4.3. Section 4.4 demonstrates theoretically that time standard sequences capture the degree of deviation from stationarity, and the proneness to choice anomalies. Section 4.5 illustrates some applications, and Section 4.6 shows how discount functions can be measured using time standard sequences. Experimental details are in Section 4.7, and results in Section 4.8. Section 4.9 contains a discussion.

4.2 Discounted utility

We consider preferences between outcome streams. An *outcome stream* $(t_1{:}x_1,\ldots,t_m{:}x_m)$ yields *outcome* x_i at *time point* t_i for $i = 1, \ldots, m$ and nothing at other time points. For simplicity of presentation we assume that outcomes are monetary and nonnegative, with "nothing" equated with the 0 outcome. Our measurement method can equally well be applied to other outcomes, with the outcome set for instance a finite set of qualitative health states, but we will not pursue this point. Time point t = 0 corresponds with the present. Under *discounted utility* (which in this paper refers to general, possibly nonconstant, discounting), outcome streams are evaluated through

$$DU(t_1:x_1,\ldots,t_m:x_m) = \sum_{i=1}^{m} \varphi(t_i)U(x_i), \tag{1}$$

where φ is the *discount function* and U the (instant) *utility function*, with $\varphi(t) > 0$ for all t, φ strictly decreasing (*impatience*) and continuous, U(0) = 0, and U continuous and strictly increasing. φ and U are *ratio scales*, meaning that each is unique up to a positive scale factor. Throughout this paper we assume that discounted utility holds. In the literature, a normalization $\varphi(0) = 1$ is sometimes assumed, but it is more convenient for us not to commit to a scaling.

Constant discounting holds if $\varphi(t) = \delta^t$ for a *discount factor* δ with $0 < \delta \leq 1$. Constant discounting has been the traditional assumption. Then a preference between two outcome streams does not change if all outcomes are delayed by an equal amount of time ε, a preference condition known as *stationarity* or *constant impatience*. Under such a delay, the discounted utility of both sequences is multiplied by the same constant δ^ε, so that their ordering is not affected. It is well-known that the reversed implication also holds under common assumptions, that is, constant impatience implies constant discounting (Koopmans, 1960).

In psychological studies it has often been found that people have *decreasing impatience*, i.e. their willingness to wait increases as outcomes are delayed. A popular model that captures decreasing impatience is the *quasi-hyperbolic* *discounting* model (Phelps and Pollak, 1968), where the discount function is given by

$$\varphi(t) = 1 \text{ if } t = 0 \text{ and}$$
$$\varphi(t) = \beta\delta^t \text{ if } t > 0, \tag{2}$$

for a constant $\beta \leq 1$ with, again, $0 < \delta \leq 1$. Under quasi-hyperbolic discounting we have decreasing impatience only at time point 0, and constant impatience thereafter.

A model that captures decreasing impatience not only for the present, but also for future time points, is *generalized hyperbolic discounting* (Loewenstein and Prelec, 1992), defined by

$$\varphi(t) = (1+gt)^{-h/g}, \tag{3}$$

with $g > 0$ and $h > 0$. Harvey (1986, Eq. 7) considered the special case $g = 1$. Herrnstein (1981) and Harvey (1995, "proportional discounting") considered the case $g = h$.

In general, violations of stationarity need not imply time inconsistency, contrary to claims sometimes made in the literature, but in agreement with some careful discussions (Dasgupta and Maskin, 2005, section I; Harvey, 1995 p. 389; Thaler, 1981). For example, you may have a special preference for apples on Tuesday, and prefer two apples on Tuesday to one apple on Monday, but not prefer two apples on Wednesday to one apple on Tuesday. This entails a violation of stationarity, but no inconsistency. All the time you consistently have and predict your preferences, and you never change plans. Such discrepancies between stationarity and time inconsistency are caused by nonhomogeneity of time. As is common in the literature, we assume homogeneous time henceforth, so that at every time point your decisions can be based on stopwatch time, and nonconstant impatience can be equated with time inconsistency. Then nonconstant impatience entails a vulnerability to arbitrage, where a person first pays to obtain an exchange one way and later pays again to reverse the exchange, ending up in the original position less some money. We will come back to this point in Section 4.4.

4.3 Deriving the degree of time inconsistency from time standard sequences

A *time standard sequence* is a sequence $t_0,\ldots,t_n$ of time points such that there exist two outcomes $\beta < \gamma$ with

$$\begin{array}{lcl} (t_0{:}\ \beta) & \sim & (t_1{:}\ \gamma) \\ (t_1{:}\ \beta) & \sim & (t_2{:}\ \gamma) \\ & \vdots & \\ (t_{n-1}{:}\ \beta) & \sim & (t_n{:}\ \gamma) \end{array} \tag{4}$$

That is, each delay between two consecutive time points exactly offsets the same improvement of outcome. This delay between two consecutive time points, $d_i = t_i - t_{i-1}$, is called the *willingness to wait* (*WTW*). Stationarity means that the WTW is constant. Under decreasing impatience the WTW increases as time proceeds, and under increasing impatience it decreases. Thus, a time standard sequence readily identifies constant, increasing, or decreasing impatience.

Time standard sequences are equally spaced in terms of the logarithm of the discount function. Because the derivation of this result may be clarifying, we give it in the main text.

OBSERVATION 1. For a time standard sequence $t_0,\ldots,t_n$:

$ln(\varphi(t_0)) - ln(\varphi(t_1)) = ln(\varphi(t_1)) - ln(\varphi(t_2)) = \cdots = ln(\varphi(t_{n-1})) - ln(\varphi(t_n))$.

PROOF. For a time standard sequence we have

$\varphi(t_0)U(\beta) = \varphi(t_1)U(\gamma)$ and

$\varphi(t_1)U(\beta) = \varphi(t_2)U(\gamma)$,

implying

$$\frac{\varphi(t_0)}{\varphi(t_1)} = \frac{U(\gamma)}{U(\beta)} = \frac{\varphi(t_1)}{\varphi(t_2)} = \cdots = \frac{\varphi(t_{n-1})}{\varphi(t_n)}. \tag{5}$$

Here the third and following equalities result from analogous algebraic manipulations. Taking logarithms gives the observation.

The points $t_0,\ldots,t_n$ are, obviously, also equally spaced in terms of normalizations of $ln(\varphi(t))$, such as at t_0 and t_n. The latter normalization is denoted τ, and is called the *time standard sequence curve*. It is given by

$$\tau(t) = \frac{ln(\varphi(t)) - ln(\varphi(t_n))}{ln(\varphi(t_0)) - ln(\varphi(t_n))} \tag{6}$$

Because it is 1 at t_0 and 0 at t_n, with n equally big steps of size 1/n in between, we get

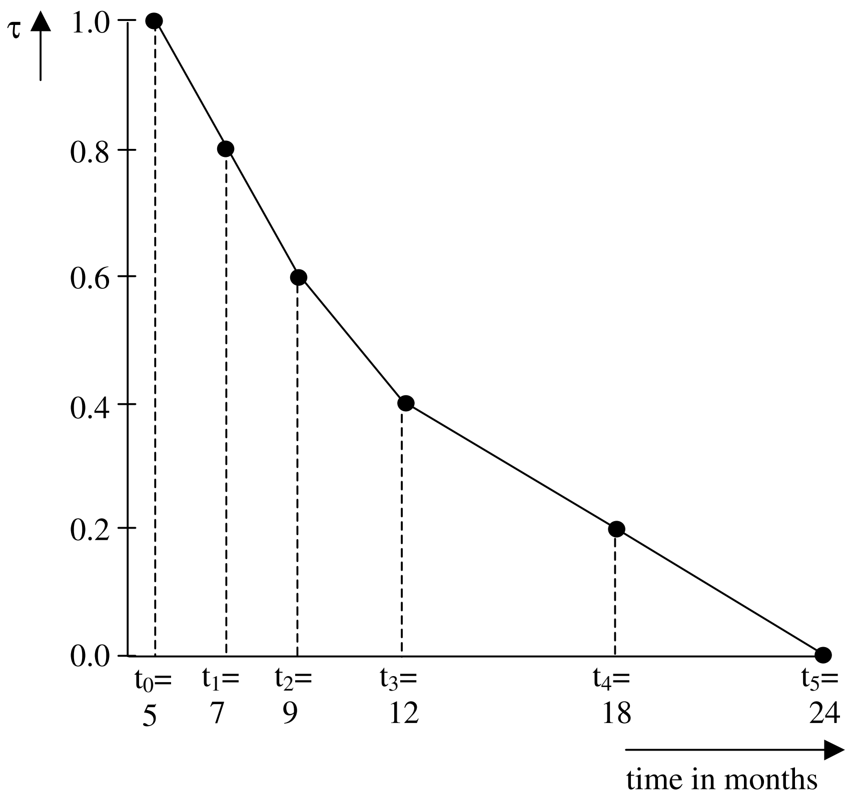
$$\tau(t_j) = 1 - \frac{j}{n} \text{ for all j.} \tag{7}$$

From time standard sequences we can, thus, immediately obtain the graph of the normalized logarithmic discount function. See Figure 4.1, with points $(t_j, 1 - \frac{j}{n})$ depicted, and linear interpolation. The figure concerns the experiment reported later, and is derived from participant 7's indifferences

(05 months: €700) ~ (07 months: €900)
(07 months: €700) ~ (09 months: €900)
(09 months: €700) ~ (12 months: €900)
(12 months: €700) ~ (18 months: €900)
(18 months: €700) ~ (24 months: €900)

so that n = 5, $t_0 = 5$, $t_1 = 7$, $t_2 = 9$, $t_3 = 12$, $t_4 = 18$, and $t_5 = 24$.

FIGURE 4.1. THE TIME STANDARD SEQUENCE CURVE OF PARTICIPANT 7.



The degree of convexity of a function is not affected by normalizations and, hence, the convexity of τ equals the convexity of $ln(\varphi)$. As we saw, stationarity, decreasing impatience, and increasing impatience correspond with constant, increasing, and decreasing WTW. Hence, we obtain the following result.

OBSERVATION 2. Stationarity corresponds with linearity of the time standard sequence curve and of $ln(\varphi)$. Decreasing impatience corresponds with convexity of the time standard sequence curve and of $ln(\varphi)$. Increasing impatience corresponds with concavity of the time standard sequence curve and of $ln(\varphi)$. The time standard sequence curve τ and the logarithm of discounting $ln(\varphi)$ have the same degree of convexity, i.e., $-\frac{\tau''}{\tau'} = -\frac{ln(\varphi)''}{ln(\varphi)'}$.

4.4 Time standard sequences to measure proneness to arbitrage

In this section, we restrict our attention to simple outcome streams. The following analysis will, contrary to the rest of this paper, essentially use continuity of utility. A *simple outcome stream* has at most one nonzero outcome, and can be written as $(s{:}\alpha)$. Consider the following two indifferences, similar to expression (4):

$(s{:}\beta) \sim (t{:}\gamma)$ and $(s+\sigma{:}\beta) \sim (t+\sigma+\rho{:}\gamma)$
for $s < t$ (s for "short"), $\beta < \gamma$, and $\sigma > 0$. (8)

We have $\rho > 0$ under decreasing impatience, $\rho = 0$ under constant impatience, and $\rho < 0$ under increasing impatience. ρ can be taken as an index of deviation from stationarity. Indeed, for $\rho > 0$, we have the typical nonstationarity

$(s:\beta) \succcurlyeq (t':\gamma)$ and $(s+\sigma:\beta) \preccurlyeq (t'+\sigma:\gamma)$ with one preference strict (9)

for all $t \leq t' \leq t+\rho$ and for no other t'. The interval $[t, t+\rho]$, thus, indicates a space for arbitrage.

Preference reversals as in (9) are prone to arbitrage. At time 0 the person, when endowed with $(s+\sigma:\beta)$, is willing to exchange it for $(t'+\sigma:\gamma)$. When asked to reconsider at time point σ, the person now perceives of the options as $(s:\beta)$ and $(t':\gamma)$, and is willing to go back to the β-option.[2] The person is willing to pay a small amount for at least one of the two exchanges (take it small enough not to affect preference otherwise). Then the person has ended up at the original endowment less some money, which entails arbitrage. Bénabou and Tirole (2002), Gruber and Köszegi (2001), Laibson (1997), O'Donoghue and Rabin (1999), Prelec (2004), Strotz (1956), Thaler and Benartzi (2004), and numerous others derived various choice anomalies from (9), and gave formalizations for these phenomena. For example, a sophisticated person who is informed about the above procedure beforehand may avoid it but then becomes vulnerable to commitments to dominated options, due to lack of future self-control. Other anomalies that can result entail time inconsistency, addiction, and procrastination.

For $\rho < 0$ in (8), as typical of increasing impatience, we have

$(s:\beta) \preccurlyeq (t':\gamma)$ and $(s+\sigma:\beta) \succcurlyeq (t'+\sigma:\gamma)$ with one preference strict (10)

---

[2] In the latter step we use homogeneity of time, i.e. the possibility to use stopwatch time, as assumed throughout this chapter.

for all $t+\rho \le t' \le t$ and for no other t', and $[t+\rho, t]$ indicates a space for arbitrage.

Consider now another preference relation $\succcurlyeq^*$, satisfying the assumptions of preceding sections as does $\succcurlyeq$, with corresponding φ^*, U^*, τ^*.

DEFINITION 1. $\succcurlyeq^*$ exhibits *more decreasing impatience* than $\succcurlyeq$ if the equivalences in (8) plus $(s{:}\beta^*) \sim^* (t{:}\gamma^*)$ imply $(s+\sigma{:}\beta^*) \preccurlyeq^* (t+\sigma+\rho{:}\gamma^*)$.

Prelec (2004) gave an equivalent definition. Under decreasing impatience for $\succcurlyeq$ and $\succcurlyeq^*$, the above condition implies, for

$$(s{:}\beta^*) \sim^* (t{:}\gamma^*) \text{ and } (s+\sigma{:}\beta^*) \sim^* (t+\sigma+\rho^*{:}\gamma^*), \tag{11}$$

that either this ρ^* exceeds ρ, or that such a ρ^* does not exist. In the first case the space $[t, t+\rho^*]$ for arbitrage for $\succcurlyeq^*$ exceeds the corresponding space $[t, t+\rho]$ for $\succcurlyeq$. In the second case the space for arbitrage for $\succcurlyeq^*$ is in fact $[t, t+\infty)$, as is readily verified.

There is also interest in increasing impatience, because of which we extend the above definition.

DEFINITION 2. $\succcurlyeq^*$ exhibits *more increasing impatience* than $\succcurlyeq$ if the equivalences in (8) plus $(s{:}\beta^*) \sim^* (t{:}\gamma^*)$ imply $(s+\sigma{:}\beta^*) \succcurlyeq (t+\sigma+\rho{:}\gamma^*)$.

For preference relations with increasing impatience, the arbitrage space [t+ρ, t] is bigger as increasing impatience is bigger.

The following theorem shows that time standard sequence curves identify proneness to arbitrage in the above sense. As usual, τ* is *more concave* than τ if there exists a concave transformation f such that τ*(t) = f(τ(t)) for all t, which holds if and only if $\frac{\tau^{*\prime\prime}}{\tau^{*\prime}} \geq \frac{\tau^{\prime\prime}}{\tau^{\prime}}$ everywhere on their domain. Note here that τ and τ* are decreasing functions, for which the Pratt-Arrow index of concavity drops the minus sign relative to increasing functions. Similarly, τ* is more convex than τ if there exists a convex transformation f such that τ*(t) = f(τ(t)) for all t, which holds if and only if $-\frac{\tau^{*\prime\prime}}{\tau^{*\prime}} \geq -\frac{\tau^{\prime\prime}}{\tau^{\prime}}$ everywhere on their domain. The following theorem adapts Prelec's (2004) Proposition 1 to time standard sequence curves instead of *ln*(φ), and extends the result to increasing impatience.

THEOREM 1. Assume that ≽ and ≽* satisfy the assumptions of discounted utility of this chapter, with ≽'s time standard sequence curve τ(t) a normalization $\frac{ln\varphi(t) - ln\varphi(S)}{ln\varphi(T) - ln\varphi(S)}$ of *ln*(φ(t)) and ≽*'s time standard sequence curve τ*(t) a normalization $\frac{ln\varphi^*(t) - ln\varphi^*(S^*)}{ln\varphi^*(T^*) - ln\varphi^*(S^*)}$ of *ln*(φ*(t)), for some arbitrary S > T and S* > T*.

(i) $\succcurlyeq^*$ exhibits more decreasing impatience than $\succcurlyeq$ if and only if $\succcurlyeq^*$'s time standard sequence curve τ^* is more convex than $\succcurlyeq$'s time standard sequence curve τ.

(ii) $\succcurlyeq^*$ exhibits more increasing impatience than $\succcurlyeq^*$ if and only if $\succcurlyeq^*$'s time standard sequence curve τ^* is more concave than $\succcurlyeq$'s time standard sequence curve τ.

The above theorem holds irrespective of the normalization parameters S, T, S*, T* chosen. The theorem demonstrates formally that the degree of convexity of a time standard sequence curve determines the degree of decreasing impatience and, thus, the space for arbitrage and the proneness to anomalies as discussed by Prelec (2004) and others. From a mathematical perspective, our reformulation in terms of time standard sequence curves, i.e. normalized $ln(\varphi)$ curves, may seem to be only more complex than Prelec's formulations directly in terms of $ln(\varphi)$ itself. This reformulation is, however, the essential step for obtaining the empirical status of the result. The normalized curve is directly observable whereas the nonnormalized curve is not.[3]

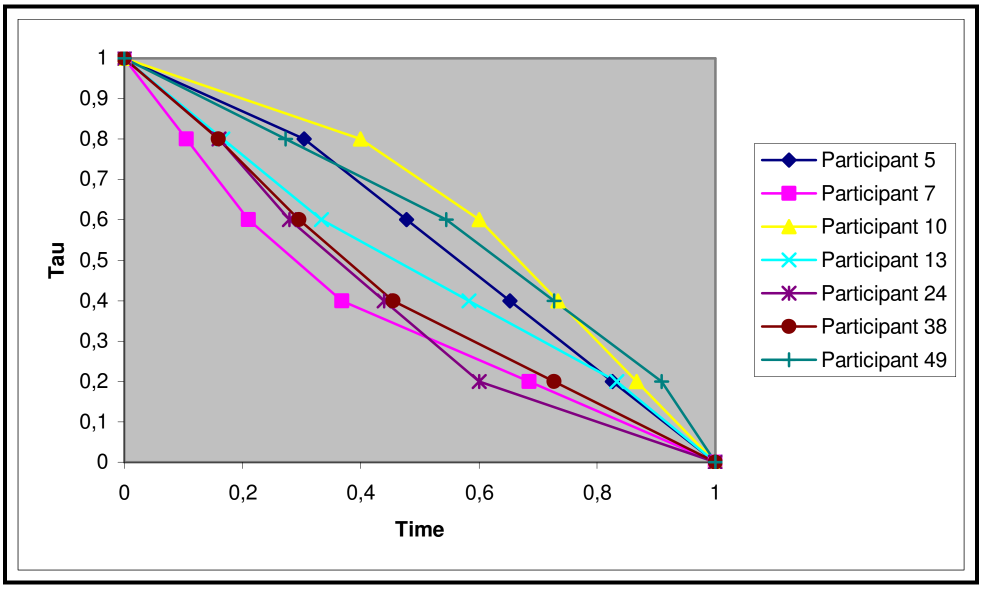
[3] In terms of Eq. (17) hereafter, Theorem 1 shows that we need not measure h, a value needed to obtain $ln(\varphi)$ but not to obtain its normalizations.

4.5 Illustrations of time standard sequences

We can immediately infer proneness to time inconsistency from simply eyeballing time standard sequence curves, curves that were obtained using only pencil and paper. Figure 4.2 displays seven time standard sequence curves, obtained from seven participants in the experiment described later, on normalized time intervals ($\tilde{t}_0$=0, $\tilde{t}_n$=1). The curves immediately reveal that the curve of participant 7 is more convex, implying more decreasing impatience, than the curve of participant 38. Because both participants exhibit decreasing impatience by Theorem 1i, participant 7 is more prone to time inconsistency and arbitrage than participant 38. Participant 24's curve is also always below that of participant 38, suggesting more decreasing impatience. Locally around 0.45, participant 38 exhibits more convexity though, so that this ordering of convexity does not hold on the whole interval [t_0,t_5]. The curves of participants 7 and 24 intersect and there is no uniform ordering regarding their degree of nonstationarity over the whole interval [t_0,t_5].

There are several concave curves suggesting increasing rather than decreasing impatience. Theorem 1ii shows that participant 10 is more prone to time inconsistency than participant 5; etc.

FIGURE 4.2. THE TIME STANDARD SEQUENCE CURVES OF SEVERAL PARTICIPANTS.



For sophisticated analyses, we can estimate ratios of second derivatives by first derivatives, or find best-fitting parametric curves, and compare the corresponding degrees of convexity. We can also develop global heuristic measures of convexity that can be calculated using only pencil and paper. For example, the area below the diagonal is a plausible index of convexity and of decreasing impatience. This area is a monotonic transform of the *decreasing-impatience index* (*DI-index*), defined by

$$\text{DI-index} = \sum_{i=1}^{n-1} \left(\frac{i}{n} - \tilde{t}_i\right), \text{ with } \tilde{t}_i \text{ the normalization of } t_i. \tag{12}$$

These values are 0.63 (participant 7), 0.52 (participant 24), and 0.36 (participant 38). They suggest that, overall, participant 7 exhibits more decreasing impatience than participant 24, and participant 24 more than participant 38. Notice that the DI-index bears some resemblance with the Gini-index in inequality measurement.

Participants 5, 10, and 49 exhibit increasing impatience. Accordingly, their DI-indices will be negative, and they are −0.26 (participant 5), −0.60 (participant 10), and −0.45 (participant 49). Overall, participant 10 exhibits more increasing impatience than participant 49, and participant 49 does so more than participant 5.

The DI-index of participant 13 is 0.08, and this participant virtually exhibits no decreasing or increasing impatience in an overall sense. Yet, this participant does deviate considerably from stationarity. For deviations from stationarity, absolute values of deviations from linearity are more relevant, with any area between the τ curve and the diagonal taken positively. We define the *non-stationarity index* (*NS-index*) as

$$\text{NS-index} = \sum_{i=1}^{n-1} \left| \frac{i}{n} - \tilde{t}_i \right| . \tag{13}$$

It provides an overall index of deviation from stationarity and proneness to inconsistencies without concern of the direction of deviation. For participant 13, the NS-index = 0.15. To the extent that stationarity is rational, the NS-index could be interpreted as an index of rationality.

The DI-index and NS-index depend on the size of the interval $[t_0,t_5]$ considered in the sense that they will tend to zero if the interval $[t_0,t_5]$ becomes small. The desirability of such dependence depends on the application considered.

Distortions due to this effect can be avoided by comparing participants only on same time intervals, or taking subparts of the τ curve related to the same interval. In Figure 4.2, the curves of participants 13 and 49 concerned similar time intervals indeed (being [5,17] and [5,16]).

Other measures of decreasing impatience or absolute deviation from stationarity can be devised, depending on the application and the time discounting assumed. Let us consider *generalized hyperbolic discounting*, $\varphi(t) = (1+gt)^{-h/g}$ (Loewenstein and Prelec 1992), with $0 \leq g < \infty$ an index of decreasing impatience, and with stationarity and constant discounting e^{-ht} the limiting case of $g \rightarrow 0$. This family incorporates most of the popular hyperbolic families other than quasi-hyperbolic discounting, such as those of Herrnstein (1981) and Harvey (1986, 1995). Rohde (2005) derived the most appropriate index of convexity for this family, called the hyperbolic factor. For a time standard sequence $t_0,\ldots,t_n$, hyperbolic factors can be calculated as

$$\textit{hyperbolic factor}(i,j) = \frac{(t_j - t_i) - (t_{j-1} - t_{i-1})}{t_i(t_{j-1} - t_{i-1}) - t_{i-1}(t_j - t_i)} \tag{14}$$

for $j > i$. Hence, for one time standard sequence with $n = 5$, 10 (= 4+3+2+1) hyperbolic factors can be calculated. As Rohde (2005) demonstrated, for generalized hyperbolic discounting, $\varphi(t) = (1+gt)^{-h/g}$, the hyperbolic factor is constant, independent of i and j or the time standard sequence considered, and is always equal to g. For constant discounting (stationarity), the hyperbolic factor will always be zero, and for quasi-hyperbolic discounting the hyperbolic factor is positive at time point 0 and zero for all future time points. Thus, this statistic can readily serve to test these models.

The relation between Prelec's convexity index and the hyperbolic factor can be compared with indices of risk aversion of utility functions U in expected utility. The absolute Pratt-Arrow index $-U''/U'$ is most appropriate for so-called CARA utility, but the relative index $x(-U''/U')$ is most appropriate for so-called CRRA utility. Thus, what is the most useful index depends on the application mentioned. Theorem 1 described cases where Prelec's measure is most suited. If we restrict our attention to generalized hyperbolic discounting, the hyperbolic factor is useful.

The hyperbolic factor can be directly calculated from time standard sequences, and can be used to test whether generalized hyperbolic discounting holds and, if it does, to distinguish between its various subfamilies. One necessary condition for generalized hyperbolic discounting to hold, and for g to be well behaved, is that the denominator in Eq. (14) be positive, i.e.

$$t_i(t_{j-1}-t_{i-1}) - t_{i-1}(t_j-t_i) > 0 \text{ for } j > i. \tag{15}$$

This inequality provides a test of generalized hyperbolic discounting, as does constantness of the hyperbolic factor when defined.

4.6 Deriving the discount function from time standard sequences

We saw in preceding sections that time standard sequences give the normalized logarithm of the discount function, and they give the degree of change of impatience and discounting through the degree of convexity of the function obtained. We did not derive the complete discount function in the preceding section, because we did not establish the rate of time preference in any absolute sense. Deriving the complete discount function from time standard sequences is the purpose of this section. One way to identify the discount function is to derive the utility function from some extra information or from some extra assumption, such as linearity as is often done in the literature. Then we can use (5) and we readily get φ.

An alternative route that does not need any assumption about utility is as follows. We can take any indifference between outcome streams with two nonzero outcomes:

$$(b{:}\gamma,\ c{:}\gamma) \sim (a{:}\gamma,\ d{:}\gamma) \text{ for } \gamma > 0 \text{ and } a < b < c < d. \tag{16}$$

We give the proof of the following observation in the main text because it demonstrates how the discount function can be calculated from (16) together with time standard sequences.

OBSERVATION 3. Given the function τ, the discount function φ is uniquely determined through one observed indifference (16).

PROOF. Let $t_n > t_0$, and assume that

$$\tau(t) = \frac{ln(\varphi(t)) - ln(\varphi(t_n))}{ln(\varphi(t_0)) - ln(\varphi(t_n))}.$$

That is, τ is a normalization of $ln(\varphi(t))$. There exist, as yet unknown, parameters λ and r such that $ln(\varphi(t)) = \lambda + h\tau(t)$, i.e.[4]

$$\varphi(t) = e^{\lambda} \times (e^{\tau(t)})^{h}. \tag{17}$$

The parameter e^{λ} is an arbitrary scaling constant without empirical implications. We may as well assume that it is e^{-h}. The power h determines the rate of time preference and is empirically relevant. For example, if we let the irrelevant factor e^{λ} be e^{-h} and rewrite $\varphi(t)$ as $e^{-h} \times (e^{h})^{\tau(t)}$, then for the special case of constant discounting and linear $\tau(t) = 1-t$, e^{-h} is the discount factor.

Time standard sequences in isolation cannot identify the power h and, hence, cannot identify the absolute rate of time preference. To see this point, note that time standard sequences concern simple outcome streams. If $\varphi(t)U(x)$ represents preferences over simple outcome streams (t:x), then so does $\varphi(t)^{h}U(x)^{h}$ for every $h > 0$, because $\varphi(t)U(x) \geq \varphi(s)U(y)$ if and only if $\varphi(t)^{h}U(x)^{h} \geq \varphi(s)^{h}U(y)^{h}$. Hence, without any assumption about utility, simple outcome streams and time standard sequences cannot identify the power of time discounting and the absolute degree of discounting, indeed.

The indifference in (16) implies that $\varphi(b)U(\gamma) + \varphi(c)U(\gamma) = \varphi(a)U(\gamma) + \varphi(d)U(\gamma)$, or $\varphi(b) + \varphi(c) = \varphi(a) + \varphi(d)$. Substituting (17) and dropping e^{λ} gives

$$(e^{\tau(b)})^{h} + (e^{\tau(c)})^{h} = (e^{\tau(a)})^{h} + (e^{\tau(d)})^{h}. \tag{18}$$

[4] We have $\lambda = ln(\varphi(t_n))$ and $h = ln(\varphi(t_0)) - ln(\varphi(t_n))$, with $\varphi(t_0)$ and $\varphi(t_n)$ unknown.

It is well-known that for all quadruples a´ ($e^{\tau(a)}$ above), b´ ($e^{\tau(b)}$ above), c´ ($e^{\tau(c)}$ above), and d´ ($e^{\tau(d)}$ above) with a´ > b´ > c´ > d´ (recall that τ is decreasing) there exists a unique real number h such that exactly one of the following equations holds:

$$b'^h + c'^h = a'^h + d'^h \text{ with } h > 0 \tag{19}$$

$$ln(b') + ln(c') = ln(a') + ln(d'), \text{ corresponding to } h = 0 \tag{20}$$

$$b'^h + c'^h = a'^h + d'^h \text{ with } h < 0 \tag{21}$$

Such equations have, for instance, been studied in decision under risk with expected utility where (b´,c´) and (a´,d´) designate fifty-fifty lotteries for money, and constant relative risk averse utility $U(x) = x^h/h$ ($U(x) = ln(x)$ for h = 0) is used to fit data. Unfortunately, there is no analytic expression for the solution h to the best of our knowledge, but h can readily be determined numerically.

In our above analysis for time preference, only positive powers h (= $ln(\varphi(t_0)) - ln(\varphi(t_n))$) are possible. In the experiment we measured (16) empirically, and then solved for h. If negative h resulted, it followed that the discounted utility model was falsified. To clarify how such a violation can arise, assume a time standard sequence $t_0,\ldots,t_5$ (n = 5). Assume that we take a = t_1, b = t_2, and c = t_3, in (16), and the participant chooses d < t_4. Then $\varphi(t_2) + \varphi(t_3) = \varphi(t_1) + \varphi(d) > \varphi(t_1) + \varphi(t_4)$, so that $\varphi(t_1) - \varphi(t_2) < \varphi(t_3) - \varphi(t_4)$. This contradicts the equality $\varphi(t_3) - \varphi(t_4) = \mu^2(\varphi(t_1) - \varphi(t_2))$ for $0 < \mu = \varphi(t_2)/\varphi(t_0) < 1$, and the general discounted utility model has been falsified.

Figure 4.3 depicts a discount function that we obtained for participant 5. We used his indifferences (0:700) ~ (6:900), (6:700) ~ (12:900), (12:700) ~ (16:900), (16:700) ~ (20:900), (20:700) ~ (24:900), which yields the time standard sequence $t_0 = 0$, $t_1 = 6$, $t_2 = 12$, $t_3 = 16$, $t_4 = 20$, and $t_5 = 24$. Further we used his indifference (12:700, 16:700) ~ (6:700, 24:700) as a version of (16) to estimate h in

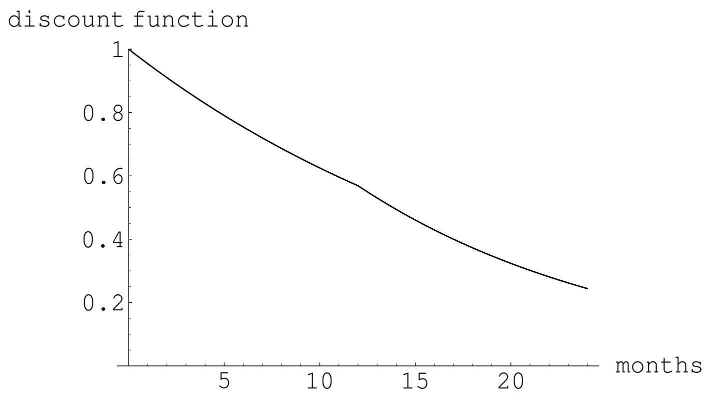
$$(e^{\tau(12)})^h + (e^{\tau(16)})^h = (e^{\tau(6)})^h + (e^{\tau(24)})^h. \tag{22}$$

Eq. (22) is equivalent to

$$(e^{3/5})^h + (e^{2/5})^h = (e^{4/5})^h + (e^0)^h. \tag{23}$$

We obtained an estimated power h = 1.41.

FIGURE 4.3. THE DISCOUNT FUNCTION φ(T) OF P5



4.7 Method of experiment

Participants

N = 55 participants took part. There were 31 students from Erasmus University, of whom 21 were from finance or economics and the others were from various other disciplines, and there were 24 students from Maastricht University, with 2 students from economics or finance and the rest from various disciplines.

Motivating participants

Every participant received €10 for participating. All payoffs in the stimuli were hypothetical. This point is discussed in Section 4.8.

Procedure

The experiment was run by computer, and participants were interviewed individually. On average, the task took 15 minutes per participant. We ran extensive pilots with 53 participants in order to determine the appropriate setup of the experiment.

We took one month as unit of time. Participants first went through a training phase, where preferences $(0{:}700) \prec (1{:}900)$ and $(0{:}700) \succ (600{:}\ 900)$ were mostly observed (with sometimes one or both reversed). Then, in a training matching task, we asked for the value t to give the indifference $(0{:}700) \sim (t{:}900)$, and then for the value t to give the indifference $(0{:}2800) \sim (t{:}3300)$.

Stimuli

We elicited four time standard sequences for each participant (Table 4.1). Every sequence consisted of 5 steps, i.e. n = 5. All tasks were matching tasks, similar to the last task of the training phase.

TABLE 4.1. PARAMETERS FOR THE 4 TIME STANDARD SEQUENCES.

Sequences	t_0	β	γ
I	0 months	€700	€900
II	0 months	€2800	€3300
III	5 months	€700	€900
IV	0 months	€1600	€1900

The outcomes β, γ, and the initial time point t_0 are as in (4).

The computer screen was as given in Appendix 4B. The pilots suggested that a direct successive elicitation of the time points $t_1,\ldots,t_5$ of one time standard sequence could generate order effects. Hence, in the main experiment we first elicited t_1 for every time standard sequence, next t_2 for every time standard sequence, etc.

We elicited two versions of (16). In both we took γ = €700. In the first we measured an indifference

$$(5{:}700,\ 11{:}700) \sim (1{:}700,\ T{:}700). \tag{24}$$

where participants were asked to provide their indifference value T through a matching question. In the second we measured an indifference

$(t_2{:}700, t_3{:}700) \sim (t_1{:}700, T{:}700)$, (25)

where t_1, t_2, and t_3 were from the elicited time standard sequence I.

Demographic variables

At the end of the experiment, participants were asked to report their gender, age, length, weight, field of study, nationality, and also whether or not they smoked. We also asked seven behavioral questions on a scale from 1 to 7, where 1 means totally disagree and 7 means totally agree. The questions concerned behavioral aspects that we thought could have an influence on discounting and are given in Appendix 4B.

Analysis

We did all tests both parametrically and nonparametrically. These always gave similar results, and we only report the nonparametric tests.

Analysis of group averages

Changes in WTW indicate whether participants satisfy constant, decreasing or increasing impatience. We tested for constant WTW for each time standard sequence separately using a Friedman test.

Next, for every two subsequent measurements of WTW (d_i and d_{i-1}) we tested equality using Wilcoxon tests. We also tested equality of WTW between the first questions of sequence I ((0:700) ~ (t:900)) and of sequence III ((5:700) ~ (t:900)). Because these concern the same outcomes, stationarity predicts the same WTW here.

We also checked whether the temporal attitude suggested by this comparison is consistent with the temporal attitude suggested by comparisons within sequence I. That is, we checked whether the change in WTW from the first question of sequence I to the first question of sequence III has the same sign as the first change in WTW within sequence I.

Analyses of individual data

A participant was classified as exhibiting increasing (constant, decreasing) impatience if at least 50% of her changes in WTW suggested so, where we considered all sequences together. A double classification as constant or increasing (decreasing) was reclassified as increasing (decreasing), and a double classification as increasing and decreasing was taken as unclassified, as were all other cases. We used these conservative criteria to reduce the effects of response error. Such a threshold of 50% has been used before in the literature (e.g. Abdellaoui, 2000). We tested whether significantly more participants are classified as increasingly or decreasingly impatient using Wilcoxon.

Next, we tested whether quasi-hyperbolic discounting holds. For every participant we split all changes in WTW of all time standard sequences into two groups: the group containing all changes in WTW where the first time point was 0, and the group containing the rest. For both groups, we chose the same 50% classification as before. Under quasi-hyperbolic discounting, the WTW should increase in the former group and be constant in the latter. We performed similar Wilcoxon tests as before.

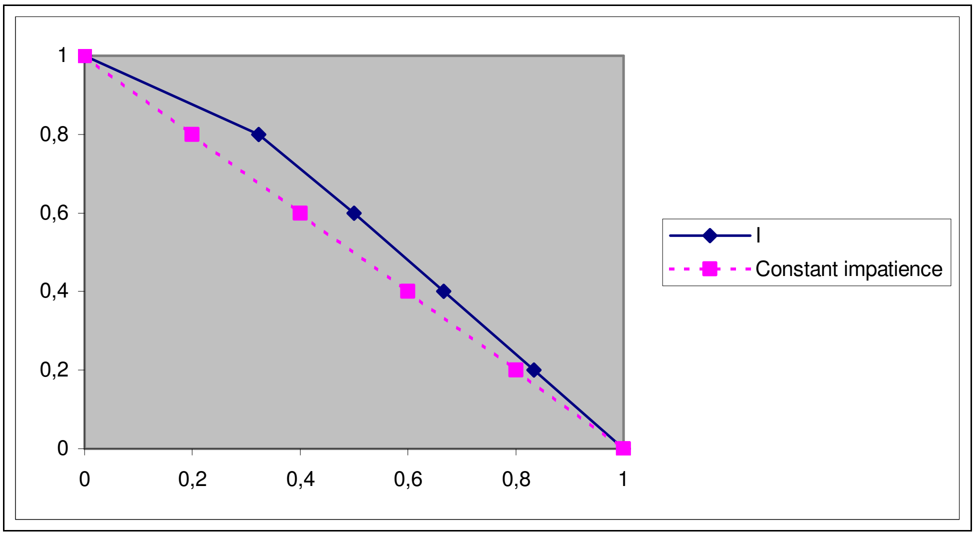
For every participant we calculated indices of decreasing impatience and of non-stationarity, and also the hyperbolic factors as explained in Section 4.5. We compared the indices of all participants between sequences by means of Wilcoxon signed rank tests. To test for a possible special effect of first questions, we also considered sequences with the first step left out. We computed the DI-index for these reduced sequences as follows: DI-index = $\sum_{i=1}^{3} \left(\frac{i}{4} - \tilde{t}_{i+1} \right)$, with $\tilde{t}_i$ the normalization of t_i such that $\tilde{t}_1 = 0$ and $\tilde{t}_5 = 1$.

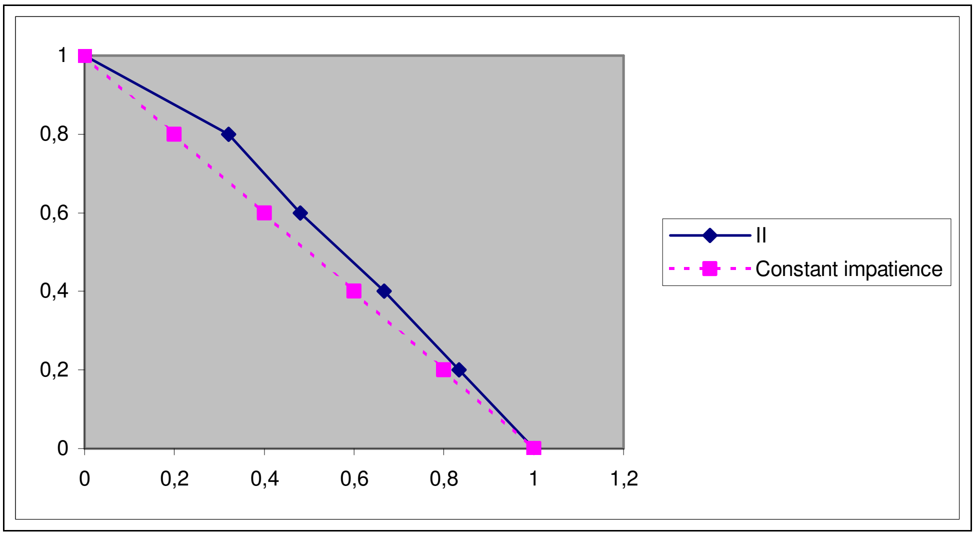
We thereafter regressed the indices of decreasing impatience and non-stationarity on the body-mass index. We estimated the correlation between each of the seven behavioral questions and each DI-index and each NS-index. We also regressed the mean of the DI-indices per participant and that of the NS-indices on gender, smoker, age, length, weight, and all behavioral questions together. Finally, we estimated the power of discounting h in Eq. (17) from the questions in (24) and (25).

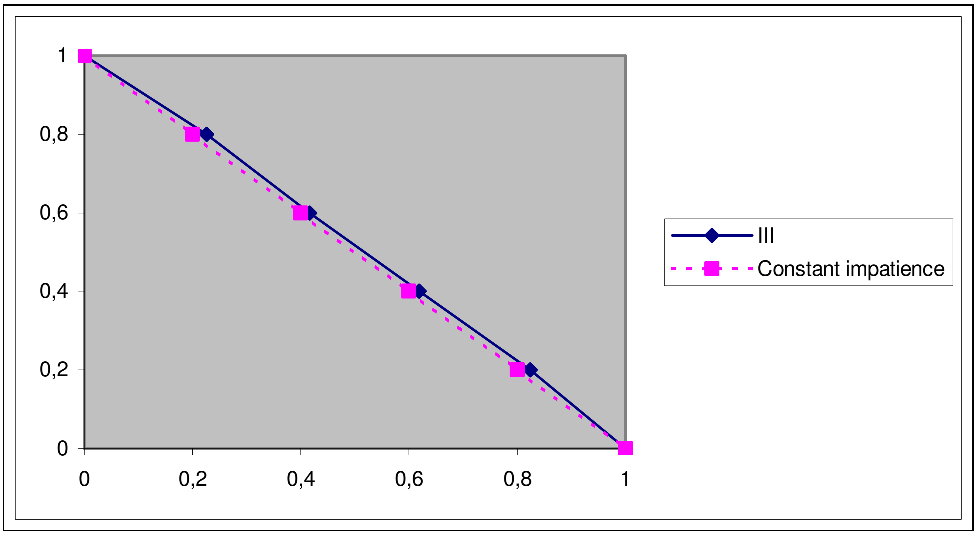
4.8 Results

Group averages

FIGURE 4.4. THE TIME STANDARD SEQUENCE CURVES FOR MEDIAN ANSWERS OF THE 4 TIME STANDARD SEQUENCES.






1
0,8
0,6
0,4
0,2
0
0
0,2
0,4
0,6
0,8
1
III
Constant impatience

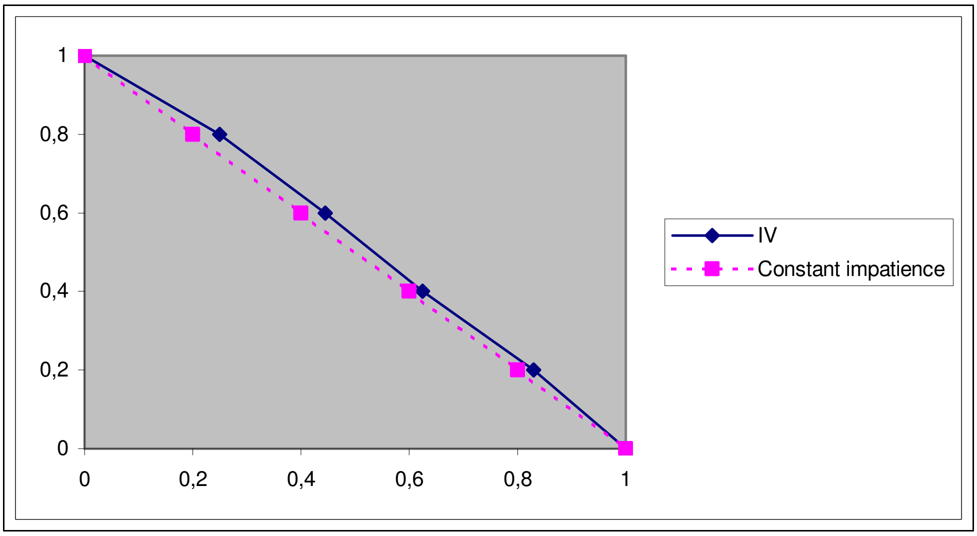
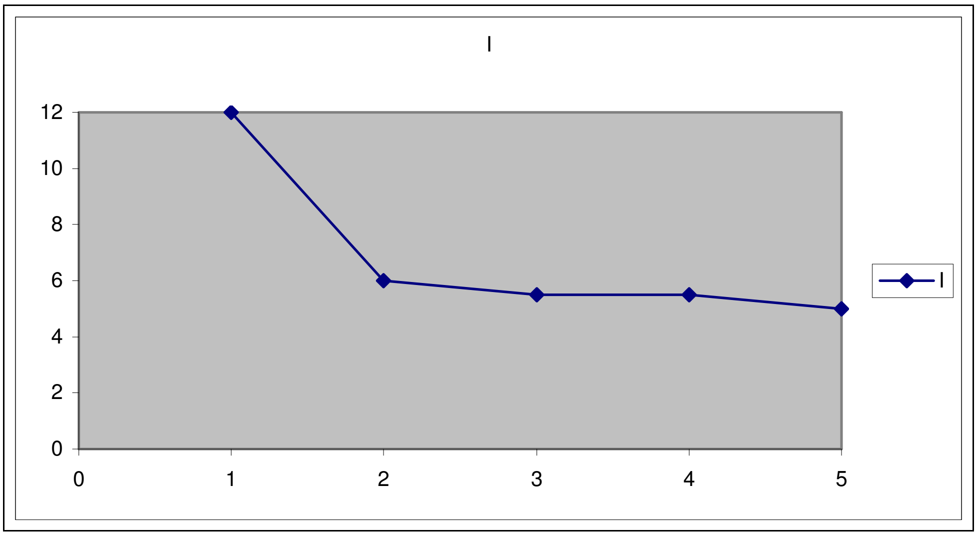
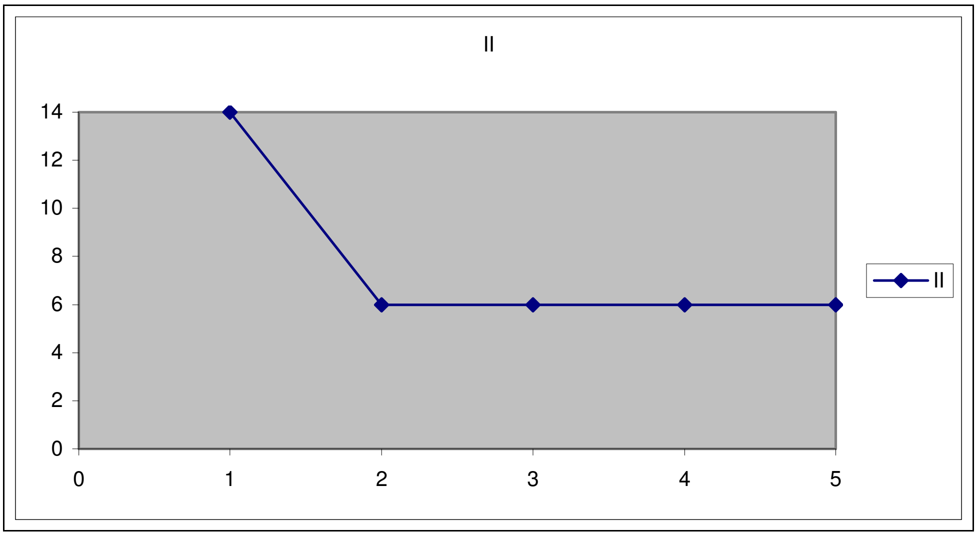
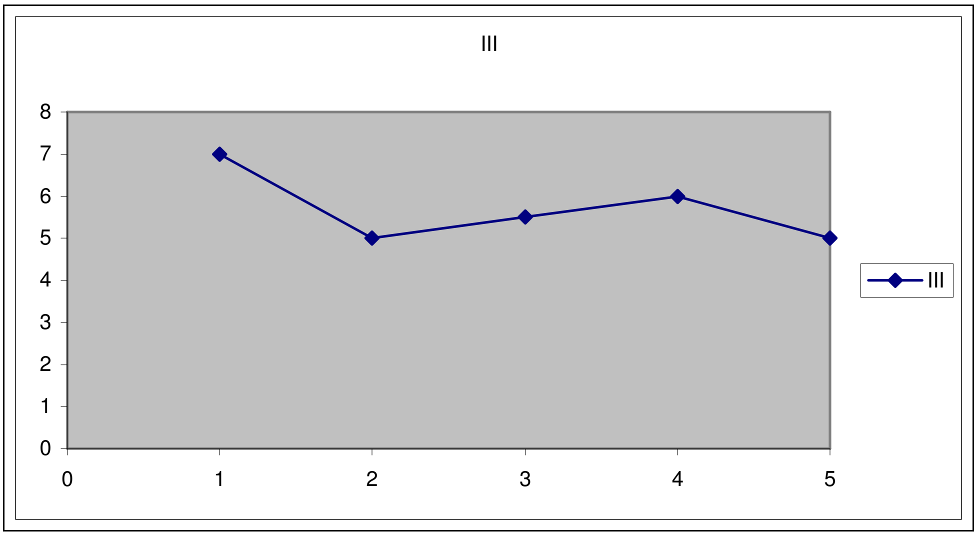

1
0,8
0,6
0,4
0,2
0
0
0,2
0,4
0,6
0,8
1
IV
Constant impatience

Figure 4.4 gives the time standard sequence curves constructed from the medians of the answers of all participants. The curves suggest that participants are increasingly impatient in the beginning and near future and constantly impatient thereafter. Statistical analyses confirm this pattern. The Friedman tests rejected constantness of the WTW ($p < 0.01$) for all sequences. We repeated the test with the first WTW excluded. As expected, then the null hypothesis of constant WTW is not rejected ($p > 0.20$ for all tests). Thus, our findings suggest that people satisfy stationarity for time points beyond a certain threshold. From the third sequence we can see that this threshold exceeds 5 months.

FIGURE 4.5. MEDIAN WILLINGNESS TO WAIT FOR EACH SEQUENCE.




II
14
12
10
8
6
4
2
0
0
1
2
3
4
5
II


III
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
III

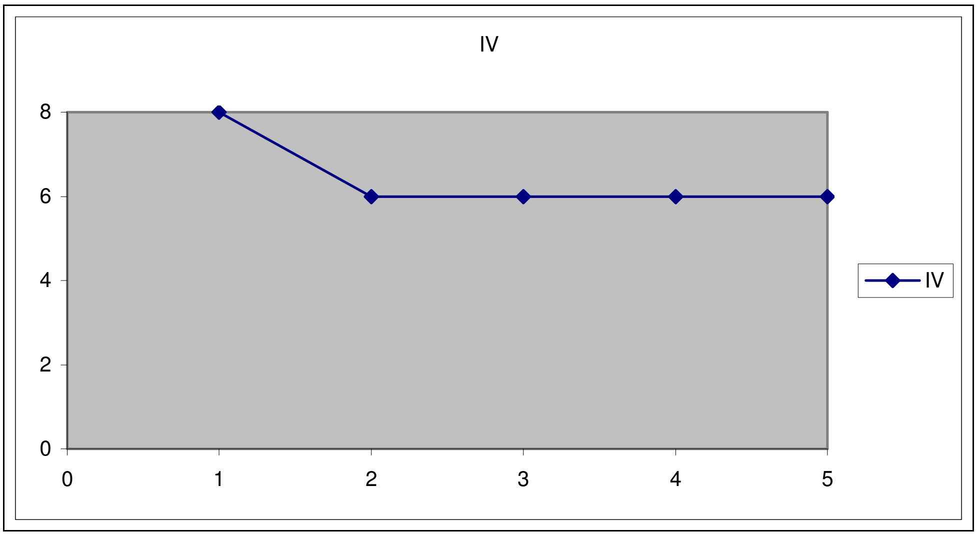


Figure 4.5 shows median WTWs. The vertical axes give the WTW. We can clearly see that the WTW drops in the beginning and remains more or less constant later on for every sequence. This is confirmed by Wilcoxon tests. The results of the Wilcoxon test are summarized in Table 4.2. The WTW changed significantly in the first steps ($d_2 - d_1$) ($\alpha = 0.01$). The WTW decreases there, suggesting increasing impatience. The WTW increases in the second step (d_3 - d_2) for sequence III ($\alpha = 0.05$). No other changes are significant at $\alpha = 0.05$.

TABLE 4.2. WILCOXON SIGNED RANK TESTS: Z (P-VALUE, 2-TAILED).

Sequence	WTW			
	$d_2 - d_1$	$d_3 - d_2$	$d_4 - d_3$	$d_5 - d_4$
I	-4.40 (0.000)	-0.51 (0.612)	-0.63 (0.531)	-0.34 (0.732)
II	-4.50 (0.000)	1.35 (0.176)	-0.93 (0.352)	-0.98 (0.329)
III	-3.39 (0.001)	2.00 (0.046)	-0.29 (0.769)	-0.95 (0.341)
IV	-3.19 (0.001)	1.03 (0.302)	-0.41 (0.681)	1.05 (0.293)

A Wilcoxon test shows that the first WTW of the third sequence is significantly lower ($p < 0.01$) than the first WTW of the first sequence. Thus, participants are consistent between sequences I and III.

Individual data

The individual data confirm the preceding findings. Participants are increasingly impatient for time points close to 0 and constantly impatient for later time points, as follows. The classification of all participants based on all sequences together yields 18 participants exhibiting constant impatience, 3 exhibiting decreasing, 10 exhibiting increasing impatience, and 24 not classified (Table 4.3). Thus, based on this classification we cannot say much about the behavior of individual participants. The Wilcoxon test shows that there is more tendency towards increasing than towards decreasing impatience ($p = 0.052$). In the group of all questions with a first time point zero, 8 participants exhibit constant impatience, 3 participants exhibit decreasing impatience, 36 participants exhibit increasing impatience, and 8 participants could not be classified. This suggests that most participants indeed are increasingly impatient for time point zero, which is supported by the Wilcoxon test (p = 0.000). In the other group (first time point positive), 21 participants exhibit constant impatience, 5 participants exhibit decreasing, 6 participants exhibit increasing, and 23 participants could not be classified. Thus, it appears that most participants indeed exhibit constant impatience for time points not too close to 0.

TABLE 4.3. CLASSIFICATION OF INDIVIDUALS.

	Impatience			
Questions	Constant	Decreasing	Increasing	Unclassified
All	18	3	10	24
Time point 0	8	3	36	8
Time point > 0	21	5	6	23

Calculations of the hyperbolic factors revealed that (15) was widely violated, for virtually all participants in many questions. This falsifies the generalized hyperbolic discounting model, and Rohde's hyperbolic factor cannot be calculated in many situations.

In view of the problems of calculating the hyperbolic factors, we only use the indices of decreasing impatience (DI-index) and of non-stationarity (NS-index) to compare participants. The medians of the indices of decreasing impatience over the whole sequences are significantly negative (p < 0.01) so that participants are increasingly impatient overall. The medians of the DI-index were respectively -0.33, -0.28, -0.092, and -0.19. We observe that the third sequence had both a lower NS-index and a lower absolute value of the DI-index. This is probably caused by the fact that the third sequence starts closer to the threshold from whereon participants satisfy constant impatience. The DI-indices of the reduced sequences, the sequences without the first steps, did not deviate significantly from zero, indicating that the increasing impatience found earlier is indeed due to the first step of every sequence.

We proceed by considering only the DI-index and NS-index of the complete sequences. Based on a Wilcoxon signed rank test, the DI-index and the NS-index are significantly different for every sequence ($p < 0.01$), where the NS-index is always larger than minus the DI-index. Since most indices of decreasing impatience are negative, this finding implies that for most participants the time standard sequence curve τ intersects the curve belonging to a linear time standard sequence curve at least once. Thus, most participants are not clearly either increasingly or decreasingly impatient, but are a mix of both.

There was no significant difference in DI-index and NS-index between sequences I and II and between sequences III and IV. For all other pairs of sequences, the sequence with the higher sequence number provided significantly higher DI-indices and significantly lower NS-indices than the ones before ($p < 0.01$ for all but one, $p < 0.05$ for all). Thus, participants became less non-stationary and more decreasingly impatient or, equivalently, less increasingly impatient in later sequences.

On average, men had higher DI-indices and lower NS-indices, except for the DI-index in sequence III, but the differences were usually not significant, with marginal significance ($p < 0.10$) for the DI-index of sequence I and for the NS-indices of sequences I and II, and significance ($p < 0.05$) only for the NS-index of sequence IV.

We found no significant relations between our indices and demographic variables otherwise. Also, the correlations between the behavioral questions and the indices were mostly insignificant, so this gives no clear indication that the behavioral questions predict behavior. In the two regressions of the mean of the DI-indices per participant and that of the NS-indices on gender, smoker, age, length, weight, and all behavioral questions together, only the coefficient on the second behavioral question with the mean of the NS-index as dependent variable, was significantly positive ($p = 0.045$) and all other coefficients were insignificant.

In calculations of the power h in Eq. (17) for (24) and (25), about 1/3 of these were negative. It shows that there are many violations of the basic model of general discounting.

4.9 Discussion

Our findings suggest that the participants satisfy increasing impatience in the beginning and constant impatience thereafter. Thus, we find a kind of "reversed quasi-hyperbolic" discounting, where impatience is constant after a certain threshold and increasing as opposed to decreasing in the beginning. Impatience, however, continues to increase up to 5 months and is not constant immediately after the present. Informal discussions with participants indicated that they understood the questions and knew what they wanted to answer. For the major finding of this study that deviates from common empirical findings in the literature, i.e. increasing instead of decreasing impatience, there was clear support from the informal discussions. Many students indicated that they did not mind a delay at first, but after a long wait they extra disliked further delays. This finding is opposite to participants' becoming more insensitive to delays, as is commonly assumed in the literature.

Our finding of increasing impatience is consistent with several other studies (Airoldi et al., 2005; Frederick, 1999; Gigliotti and Sopher, 2004; Read et al., 2005a; Read et al., 2005b; Rubinstein, 2003; Sayman and Onculer, 2005). Read et al. (2005b) found that hyperbolic discounting is only observed when time is described in delay terms as opposed to calendar time terms. Rubinstein (2003) reported three experiments that provide evidence against constant or decreasing impatience. Bommier (2005) and Dasgupta and Maskin (2005) gave theoretical reasons why increasing impatience can occur.

The setup of the experiment made it unlikely that participants noticed that the questions were chained, and that several of them together served to elicit sequences. Therefore, it is unlikely that order effects would cause the increasing impatience we found.

For the violations of general discounted utility that we found when estimating the powers in Eq. (17), it is likely that the time-separability assumption underlying general discounted utility is violated. It plays no role for the measurements of time standard sequences, and only becomes effective for two or more outcomes. This finding adds to the motivation for paying attention as much as possible to simple outcome streams, as done when measuring time standard sequence curves. Therefore the analysis of the time standard sequences minimizes the biases caused by this violation.

Participants were paid a flat fee for participating and all questions were hypothetical. There are several reasons why we did not use performance-based real incentives. First, it is administratively complicated to transfer money on the time points specified, not only for the experimenters but also for the participants. Hence, such a procedure will generate many extra biases such as through doubts on the participants' part about reliable implementations. Second, the outcomes we used were large so as to avoid participants thinking that the amount of money is trivial anyhow and not worth thinking about carefully. Then real payments make the experiment prohibitively expensive. Also, no clear evidence exists that hypothetical amounts are discounted differently than real amounts (Frederick et al., 2002). In other fields with stimuli that are not cognitively demanding similar to our study, hypothetical incentives do not seem to give qualitatively different results, although real incentives tend to reduce data variability (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001). Finally, there is no clear incentive for our participants to please the experimenter, as there can be in experiments about social behavior.

Many studies that provide evidence in favor of decreasing impatience elicit indifference values in the outcome domain. They fix two time points and one outcome and elicit a second outcome that makes the participant indifferent between the two simple outcome streams. We elicit indifference values in the time domain. Because we are interested in properties of the discount function, and not of the utility function, it is more natural to focus the participants' attention on this dimension as our questions did. Because, by construction of a time standard sequence, utilities cancel out from the equations, our method does not require richness in the outcome dimension and can, for instance, be used with qualitative health outcomes. It naturally exploits the richness in the time dimension that is available anyhow.

Scale compatibility entails that participants put more weight on the time dimension in our setup than in studies eliciting indifferences in the outcome domain. This could mean that participants discount outcomes more heavily in our setup but it need not affect the main topic of interest to us: nonconstant impatience.

Although eliciting indifferences has not been very common in the time domain, it has been used on a number of occasions, for instance by Mazur (1987). He conducted experiments with pigeons instead of humans. Green et al. (1994) did similar experiments with humans. These studies, as our study, exploited the richness of the time dimension to study temporal preference. They, however, still assumed linear utility of money.

Our findings suggest a number of new directions for intertemporal preference. Virtually all existing models, including quasi-hyperbolic discounting and generalized hyperbolic discounting, assume universal decreasing or constant impatience, and have no clear extension to allow for increasing impatience. However, even if group averages satisfy decreasing impatience, then there will still be individuals who exhibit increasing impatience, so that for any data fitting at the individual level such functions are required. For this reason we could not implement the planned test to discriminate which of the currently popular models fit the data better: none of them could at all fit data. In particular, Rohde's (2005) hyperbolic factor, in theory a good tool to empirically distinguish between various families, was not defined for many answers of virtually all participants. Also when we used time standard sequences to derive discount functions from two nonzero-outcome streams, our findings were mostly negative: we found the general discounted utility model (1) extensively violated. Hence, the development of models relaxing this assumption is also desirable.

Appendix 4A. Proofs

PROOF OF THEOREM 1. Because τ and τ^* are strictly decreasing functions, $\tau^*(t) = f(\tau(t))$ for a strictly increasing function f. Take any intervals [d,c] and [b,a] to the right of [d,c] ($b > d$ and $a > c$) in the domain of f. Then $a = \tau(s)$, $b = \tau(t)$, $c = \tau(s+\sigma)$, and $d = \tau(t+\sigma+\rho)$ for some $s < t$, $s+\sigma < t+\sigma+\rho$, $\sigma > 0$, $\sigma + \rho > 0$. Because the ranges of U and U* contain nondegenerate intervals with 0 as lower bound, there exist outcomes $\beta < \gamma$ with

$$(s:\beta) \sim (t:\gamma) \tag{A.1}$$

and outcomes $\beta^* < \gamma^*$ with

$$(s:\beta^*) \sim^* (t:\gamma^*). \tag{A.2}$$

(Here is where we crucially use continuity of utility.) Only the utility ratios $U(\beta)/U(\gamma)$ and $U^*(\beta^*)/U^*(\gamma^*)$ matter for all that follows and, hence, the particular choices of β, γ, β^*, γ^* are immaterial for all that follows.

We have equivalence of the following statements:

$a - b = c - d$;

$\tau(s) - \tau(t) = \tau(s+\sigma) - \tau(t+\sigma+\rho)$;

$ln\varphi(s) - ln\varphi(t) = ln\varphi(s+\sigma) - ln\varphi(t+\sigma+\rho)$;

$\varphi(s)/\varphi(t) = \varphi(s+\sigma)/\varphi(t+\sigma+\rho)$;

$(s+\sigma:\beta) \sim (t+\sigma+\rho:\gamma)$. (A.3)

We also have logical equivalence of the following statements:

$f(a) - f(b) \geq f(c) - f(d)$.

$\tau^*(s) - \tau^*(t) \geq \tau^*(s+\sigma) - \tau^*(t+\sigma+\rho)$;

$ln\varphi^*(s) - ln\varphi^*(t) \geq ln\varphi^*(s+\sigma) - ln\varphi^*(t+\sigma+\rho)$;

$\varphi^*(s)/\varphi^*(t) \geq \varphi^*(s+\sigma)/\varphi^*(t+\sigma+\rho)$;

$(s+\sigma:\beta^*) \preccurlyeq (t+\sigma+\rho:\gamma^*)$; (A.4)

It is well-known that f is convex if and only if for all a, b, c, d as above we have $f(a) - f(b) \geq f(c) - f(d)$. As we have just demonstrated, this is, in view of (A.1) and (A.2) and the independence of the choices $\beta,\gamma,\beta^*,\gamma^*$ above, the same as the requirement that $(s+\sigma:\beta) \sim (t+\sigma+\rho:\gamma)$ imply $(s+\sigma:\beta^*) \preccurlyeq (t+\sigma+\rho:\gamma^*)$ for all s,t,σ,ρ as above. That is, convexity of f is equivalent to more decreasing impatience for $\succcurlyeq^*$ than for $\succcurlyeq$. Reversing inequalities and weak preferences above shows that concavity of f is equivalent to more increasing impatience for $\succcurlyeq^*$ than for $\succcurlyeq$.

Appendix 4B. Display screen and behavioral questions

FIGURE 4A.1. LAYOUT OF THE COMPUTER SCREEN.

The seven **behavioral questions** were as follows:

1. I do not study regularly, but often postpone it for too long, so that the exams week is extra stressful.
2. I wish I would drink less alcohol per week than I do currently.
3. I wish I would eat less per day than I do currently.
4. I tend to postpone things.
5. I am impatient.
6. I am often late.
7. I tend to do impulsive purchases.

5 A new method for measuring the utility of life duration[1]

Summary

This chapter proposes a new method to measure the utility of life duration. A major advantage of our method, as opposed to existing methods for the measurement of the utility of life duration, is that it avoids the use of probabilities and of the outcome death. People tend to have difficulties handling probabilities, leading to biases in the elicited utilities, and the possibility of death tends to produce extreme risk aversion. Our measurements are obtained by using a choice-based procedure in contrast to other studies on risk-free utility that measured utility based on direct strength of preference judgments. To compare the elicited utility of our risk-free method to utility under risk, we also measured the utility of life duration using two familiar risky elicitation methods, the certainty equivalence method and the tradeoff method. We found the certainty equivalence method to be more concave than the risk-free method, but when we adjusted the certainty equivalence method for the violations of expected utility

---

[1] This chapter is based on Attema, Bleichrodt, and Wakker (2007).

modeled by prospect theory the difference was no longer significant. In addition, the tradeoff method, which is not prone to bias caused by probability weighting, did not differ significantly from the risk-free method. Hence, our risk-free method provides a reliable measure of utility. In addition, people indicated that they found the risk-free method significantly easier than the other two methods, thereby increasing feasibility for practical purposes.

5.1 Introduction

Knowledge of the utility for life duration is crucial in making treatment recommendations that best reflect the interests of the patient, see McNeil et al. (1978) for an illustration. The common way to measure the utility for life duration is through the certainty equivalence (CE) method. In the CE a risky treatment (p:M; m), denoting probability p of M additional life-years and probability $1 - p$ of $m < M$ additional life-years, is given and the decision maker is asked to specify the number of life years x for sure that he considers equivalent to the risky treatment. Commonly, m is set equal to immediate death. Assuming expected utility and setting the utility of M equal to 1 and the utility of m to 0 it then follows that the utility of x years is equal to p.

There are two problems with the CE method. The first problem is the use of the outcome immediate death, which tends to result in extreme risk aversion, i.e. low values of x and, consequently, high utilities for life duration. The second problem is the assumption that the CE can be evaluated by expected utility.

Evidence abounds that people do not behave according to expected utility (Starmer, 2000). Violations of expected utility in medical decision making are reported, for example, by Llewellyn-Thomas et al. (1982), Rutten-van Mölken et al. (1995), Bleichrodt (2001), and Oliver (2003). The violations of expected utility lead to an upward bias in the measured utility for life duration, i.e. to utilities for life duration that are too high (Wakker and Stiggelbout, 1995; Bleichrodt et al., 2001; Bleichrodt et al., 2007).

This chapter presents a new method to measure the utility for life duration that avoids the two problems affecting the CE. Our method does not involve the outcome death and is risk-free; hence, it is not affected by violations of expected utility. The method is entirely choice-based and does not introduce additional primitives such as strength of preference judgments.

The risk-free nature of our method raises the question whether we can expect it to yield the same utilities as measurements that are applied under risk. This question touches on a long-standing issue in economics and decision theory whether utility is context-specific or whether there exists one unifying concept of utility. The dominant viewpoint has been that there is no unifying concept of utility and that the utility that is elicited from decisions under risk can only be applied in risky decisions (e.g. Arrow, 1951; Savage, 1954; Luce and Raiffa, 1957; Fishburn, 1989). In the medical domain the assumption of one unifying concept of utility is common, however. For example, time tradeoff measurements, which are made in an intertemporal context, are commonly applied in decisions involving risk.

Dyer and Sarin (1982) proposed a way to separate under expected utility riskless strength of preference from attitude towards risk. Several studies reported evidence for this distinction in the form of a significant difference between utility obtained under risk and utility obtained under certainty (Tversky, 1967; McCord and de Neufville, 1984; Miyamoto and Eraker, 1988; Camerer, 1995; Pennings and Smidts, 2000).

All these studies assumed expected utility, however. Under expected utility, the utility function captures both a decision maker's attitude towards outcomes and his attitude towards risk. Hence, if people are not risk neutral then it is plausible that riskless and risky utility will be different. Under non-expected utility, a decision maker's risk attitude is not only determined by the shape of his utility function. For example, in prospect theory risk attitude is also determined by probability weighting and loss aversion. It has been argued that utility curvature plays a minor role in shaping a decision maker's attitude towards risk and that loss aversion is the main driving force (Rabin, 2000). If utility is not influential in determining a decision maker's attitudes towards risk then riskless and risky utility can be the same. Two studies observed evidence that the difference between riskless and risky utility disappears when corrections are made for deviations from expected utility. Stalmeier and Bezembinder (1999) found that risky and riskless utility of health states did not differ anymore when they corrected for probability weighting and framing. More recently, Abdellaoui et al. (2007) observed no significant difference between risky and risk-free utility when risky utility was corrected for probability weighting. These studies lend support to the existence of an intrinsic meaning of utility, which is relevant for risky as well as riskless applications. By comparing the results from our risk-free method with the results from two risky elicitation methods, the CE method and Wakker and

Deneffe's (1996) tradeoff (TO) method, our study will shed new light on the nature of utility.

In what follows, Section 5.2 introduces notation and background. Section 5.3 describes our new risk-free (RF) method, Section 5.4 the CE method and Section 5.5 the TO method. Section 5.6 describes an experiment in which we measured the utility for life duration by the RF method, the CE method, and the TO method. Section 5.7 presents the results of the experiment and Section 5.8 concludes.

5.2 Background

Let $h = (h_1,\dots,h_T)$ denote a *health profile*, where h_t denotes health in period $t = 1,\dots,T$ and T is the decision maker's final period of life. *Constant health profiles* give the same health state at each point in time: $h_1 = \dots = h_T$ = constant. In the RF method we will consider preferences $\succcurlyeq$ over health profiles. Preferences over health states are derived from preferences over constant health profiles. If $h_1 = \dots = h_T = \gamma$ and $h'_1 = \dots = h'_T = \beta$ then $\gamma \succcurlyeq \beta$ if and only if $h \succcurlyeq h'$. We assume throughout that health profiles $h = (h_1,\dots,h_T)$ are evaluated by

$$\sum_{t=1}^{T} \delta_t V(h_t) \qquad (1)$$

and preferences and choices correspond with this evaluation. In Eq. 1 V is a real-valued *utility function* over health states and the δ_t are period-specific decision weights. We will denote $W(x) = \sum_{t=1}^{x} \delta_t$. W(x) can be interpreted as the utility of life duration x.

In the CE and the TO methods we consider preferences $\succcurlyeq$ over *binary prospects.* The prospect (p:h; g) gives health profile h with probability p and health profile g with probability 1 − p. A prospect is *riskless* if p = 1 or if h = g. We denote by $\succcurlyeq$ the decision maker's preference relation over binary prospects. We will assume throughout that binary prospects (p:h; g) are *rank-ordered*, i.e. it is implicit in the notation that $h \succcurlyeq g$.

Expected utility holds if prospects (p:h; g) are evaluated by pU(h) + (1−p)U(g), with U a real-valued function on the set of health profiles, and preferences and choices correspond with this evaluation. We will also assume that U is of the form described in Eq. 1. It is important to emphasize, however, that the decision weights δ_t and the function V may be different under the CE and the TO as the CE and the TO measure preferences under risk. Given common scaling, the CE and the TO should give the same results when expected utility holds.

We also consider deviations from expected utility described by *prospect theory* (Kahneman and Tversky, 1979). In prospect theory, preferences depend on a *reference point* r. If a health profile h is better than the reference point then it is a gain. If it is worse than the reference point it is a loss. Decision makers are assumed to be loss averse, i.e. losses loom larger than gains. Loss aversion is captured by a loss aversion parameter λ. Finally, prospect theory assumes that people do not evaluate probabilities linearly but transform probabilities. Let w^+ denote the *probability weighting function* for gains and w^- the probability weighting function for losses. The probability weighting functions are strictly increasing and satisfy $w^+(0) = w^-(0) = 0$ and $w^+(1) = w^-(1) = 1$. A *mixed prospect* (p:h; g), h>r>g, is evaluated as

$$U(r) + w^+(p)(U(h) - U(r)) - \lambda w^-(1-p)(U(r) - U(g)). \tag{2}$$

A *gain prospect* (p:h; g), h ≽ g ≽ r is evaluated as

$$U(r) + w^+(p)(U(h) - U(r)) + (1 - w^+(p))(U(g) - U(r)). \tag{3}$$

Finally, a *loss prospect* (p:h; g), r ≽ h ≽ g is evaluated as

$$U(r) - \lambda(1-w^-(1-p))(U(r) - U(h)) - \lambda w^-(1-p))(U(r) - U(g)). \tag{4}$$

Expected utility is the special case of prospect theory where $w^+(p) = w^-(p) = p$ and $\lambda = 1$. In Eqs. 2-4 the function U is assumed to be of the form specified in Eq. 1. Again, the δ_t and V can be different from the functions that apply in the RF method. If prospect theory holds and we evaluate the CE and the TO according to prospect theory then they should give the same result.

5.3 The risk-free method

The essence of the risk-free method is to ask a decision maker when he would like to have an improvement in his health. Specify two health states β (bad) and γ (good), with γ ≻ β. The decision maker always chooses between two profiles A, which gives the health improvement from β to γ sooner and B, which gives the improvement from β to γ later. Let $\gamma_{x\to y}\beta$ denote the health profile β that gives an improvement in health from β to γ from time point x until time point y. In the first assessment health profile A gives $\gamma_{0\to x_{0.5}}\beta$, i.e. an immediate improvement in health from β to γ until time point $x_{0.5}$ and B gives $\gamma_{x_{0.5}\to T}\beta$, i.e. an improvement from β to γ from time point $x_{0.5}$ until time point T. The time point $x_{0.5}$ is varied until the respondent is indifferent between A and B. This implies by Eq. 1 that

$$W(x_{0.5})[V(\beta) + (V(\gamma)-V(\beta))] + (W(T) - W(x_{0.5}))V(\beta) =$$
$$W(x_{0.5})V(\beta) + (W(T) - W(x_{0.5}))[V(\beta) + (V(\gamma)-V(\beta))]. \tag{5}$$

Rearranging yields

$$W(T)V(\beta) + W(x_{0.5})(V(\gamma)-V(\beta)) = W(T)V(\beta) + (W(T) - W(x_{0.5}))(V(\gamma)-V(\beta)). \tag{6}$$

And thus

$$W(x_{0.5}) = ½\, W(T). \tag{7}$$

Setting $W(T) = 1$ gives $W(x_{0.5}) = 0.5$.

Once $x_{0.5}$ is known, we can now proceed to elicit the utility for life duration. For example, in a next step we determine the life duration $x_{0.25}$ that makes the profiles $A = \gamma_{0\to x_{0.25}}\beta$ and $B = \gamma_{x_{0.25}\to x_{0.5}}\beta$ equivalent. Then

$$W(x_{0.25})[V(\beta) + (V(\gamma)-V(\beta))] + (W(T) - W(x_{0.25}))V(\beta) =$$
$$W(x_{0.25})V(\beta) + (W(x_{0.5}) - W(x_{0.25}))[V(\beta) + (V(\gamma)-V(\beta))] + (W(T) - W(x_{0.5}))V(\beta), \quad (8)$$

and rearranging gives $W(x_{0.25}) = ½*W(x_{0.5}) = 0.25$. To determine the life duration that has utility 0.75, we elicit the life duration $x_{0.75}$ that establishes indifference between the profiles $A = \gamma_{0\to x_{0.75}}\beta$ and $B = \gamma_{x_{0.25}\to T}\beta$. This gives $W(x_{0.75}) = W(T) - W(x_{0.25}) = 0.75$. Continuing in this manner we can measure W without making particular assumptions about its functional form.

5.4 The certainty equivalence method

In the first step of the certainty equivalence method we elicited the sure outcome $x_{0.5}^{ce}$ that a decision maker considers equivalent to a risky prospect (½:T; 0). All years are spent in the same health state so for notational convenience we will suppress the health state. By expected utility $x_{0.5}^{ce} \sim (½{:}T; 0)$ implies $W_{CE}(x_{0.5}^{ce}) = 0.5$. We can then proceed to elicit $x_{0.25}^{ce}$ and $x_{0.75}^{ce}$ through the indifferences $x_{0.25}^{ce} \sim (½{:}x_{0.5}^{ce}; 0)$ and $x_{0.75}^{ce} \sim (½{:}T; x_{0.5}^{ce})$. It is easily verified that under expected utility $W(x_{0.25}^{ce}) = 0.25$ and $W(x_{0.75}^{ce}) = 0.75$. These outcomes can be used in turn to elicit further points of the utility for life duration.

To determine the utility for life duration under prospect theory we must know the decision maker's reference point. Bleichrodt et al. (2001) presented evidence that the reference point in the CE was equal to 0 and participants generally perceive all life durations in the CE as gains. Then it follows from Eq. 3 that

$$W_{CE}(x^{ce}_{0.5}) = w^{+}(0.5)$$
$$W_{CE}(x^{ce}_{0.25}) = w^{+}(0.25)$$
$$W_{CE}(x^{ce}_{0.75}) = w^{+}(0.5) + (1 - w^{+}(0.5))w^{+}(0.5). \tag{9}$$

5.5 The tradeoff method

In the TO method (Wakker and Deneffe, 1996) a sequence of outcomes that are equally spaced in utility units is established. As in the CE, health status is held constant and we will therefore suppress the level of health in the ensuing notation. The first step in the TO method is to select two *gauge durations* G and g (G > g), a probability p, and a *starting life duration* x^0. Then the life duration x^1 is elicited such that the decision maker is indifferent between the prospects (p:G; x^0) and (p:g; x^1). We will analyze this equivalence by prospect theory. The evaluation under expected utility is the special case where $w^+(p) = p$ and all properties of the measurements derived below also hold under expected utility. To be able to apply prospect theory we must know the location of the reference point. Here we follow Bleichrodt et al. (2001) who showed that people perceive all outcomes in the TO method as gains. Provided that $x^1 < g$, the indifference (p:G; x^0) ~ (p:g; x^1) implies that

$$w^+(p)W(G) + (1 - w^+(p))W(x^0) = w^+(p)W(g) + (1 - w^+(p))W(x^1). \tag{10}$$

Thus

$$W(x^1) - W(x^0) = \frac{w^+(p)}{1 - w^+(p)}(W(G) - W(g)). \tag{11}$$

The life duration x^1 is used as an input in the next question in which x^2 is elicited such that the decision maker is indifferent between the prospects $(p:G; x^1)$ and $(p:g; x^2)$. By (1) and a similar argument as above this implies that

$$W(x^2) - W(x^1) = \frac{w^+(p)}{1 - w^+(p)}(W(G) - W(g)). \tag{12}$$

And thus, $W(x^2) - W(x^1) = W(x^1) - W(x^0)$. We proceed to elicit indifferences $(p:G; x^{j-1}) \sim (p:g; x^j)$, $j = 1,\ldots,k$. As long as $x^j \leq g$, these indifferences imply that $W(x^j) - W(x^{j-1}) = W(x^1) - W(x^0)$, $j = 1,\ldots,k$. Hence, the TO method elicits a *standard sequence* of life duration $x^0,x^1,\ldots,x^k$, which are equally spaced in terms of utility, i.e. the utility difference between successive elements of the standard sequence is constant. Normalizing the utility of life duration such that $W(x^0) = 0$ and $W(x^k) = 1$ gives $W(x^j) = j/k$, $j = 0,\ldots,k$. If $x^j > g$, the rank ordering of the prospects is affected, Eq. 1 can no longer be applied and the above derivations do not apply.

Equations 11 and 12 show that the elicitation of the standard sequence, and, hence, the measurement of the utility for life duration, is independent of the probability weighting function. Consequently, the TO method is not prone to bias due to probability weighting. Bleichrodt et al. (2001) showed that the TO method is also robust to loss aversion in the most plausible cases. Therefore, if a significant difference is found between the results of the TO method and the results of the RF method, this is not caused by probability weighting or loss aversion.

5.6 Experiment

Participants

Seventy participants enrolled in the experiment and were paid a fixed amount of €12.50. The participants were students from different faculties of the Erasmus University Rotterdam. Before the actual experiment, we tested the design in several pilot sessions using other students and university staff as participants.

Procedure

The experiment was computer-based and was administered in sessions of at most two persons. An experimenter was present during each session. All participants finished the experimental session within 45 minutes. To avoid order effects, the order of the RF, CE, and TO methods was randomized across sessions. We did not intersperse the questions of the different methods so the three methods were administered successively. All methods were preceded by practice questions.

All indifferences were elicited through a sequence of choices between two options, which were neutrally labeled A and B. We used choices since there is empirical evidence suggesting that choice-based procedures cause fewer inconsistencies than matching procedures, in which participants are asked to state their indifference value directly (Bostic et al., 1990). The indifferences were obtained by an iterative process, which is explained in Appendix 5C. After each choice the participant was asked to confirm his choice. Indifference values were elicited in at most five iterations. At the end of the iteration process, the first choice of the process was repeated to try and minimize the impact of response error. If the respondent changed his choice the iteration process was started anew. To check for consistency, the elicitation of the first indifference value was repeated at the end of each part.

Stimuli of the RF method

Health state β was specified as regular back pain. We selected this health state because it is a common illness and the participants were likely to know people suffering from it. We described the health state by the Euroqol 5D questionnaire. Health state γ was specified as full health. It was made clear to the participants that this health state meant they were able to function perfectly on all five dimensions, irrespective of their age. The descriptions were printed on cards and handed to the participants (see Appendix 5A).

The life duration T was set equal to 50 years. In the first question we determined $x_{0.5}$ such that $W(x_{0.5}) = ½*W(50) = 0.5$. As described in Section 5.3, we did this by finding the life duration for which a participant was indifferent between $A = \gamma_{0 \to x_{0.5}} \beta$ and $B = \gamma_{x_{0.5} \to T} \beta$. We told the participants that after this period the two options gave the same health state without further specifying it. We also used the RF method to elicit $x_{0.125}$, $x_{0.25}$, $x_{0.75}$ and $x_{0.875}$.

Stimuli of the CE method

In the CE part of the experiment, participants had to compare the risk-free option x years in full health with a risky option (½: z years in full health; y years in full health), with $y < x < z$. Full health (FH) was used as the prevailing health state throughout. In the first question we set $z = T = 50$ years and y = death within a week. This made it possible to elicit $x^{ce}_{0.5}$. As in the RF method, we then used $x^{ce}_{0.5}$ as input to elicit $x^{ce}_{0.125}$, $x^{ce}_{0.25}$, $x^{ce}_{0.75}$, and $x^{ce}_{0.875}$.

Stimuli of the TO method

The gauge outcomes *G* and *g* in the TO method were set equal to 55 and 45 years and p was set equal to ½. The initial outcome x_0 was death within a week. A standard sequence $x_1,\ldots,x_6$ was determined by eliciting the value of x_j that established indifference between (½:55; x^{j-1}) and (½:45; x^j), j = 1,…,6.

Convenience of the methods

We included a question at the end of the experiment to explore whether participants preferred one of the methods to the others in terms of understandability and cognitive burden. The participants were asked to rate each method on a scale of 1 to 7, where 1 indicated that the questions of the method were very hard to answer and 7 indicated that they were very easy to answer. An example can be found in Appendix 5D.

Analyses

Because the RF and the CE were both measured on the same domain [0,50], they could be directly compared. The comparison of RF and CE with the TO is more involved because the highest outcome in the TO method, x^6, is in general different from 50 years, the highest outcome in the RF and CE elicitations. As a result, the domains of the utility functions did not coincide and we could only compare utility curvature on common subdomains. In those cases utilities were rescaled (to W^*) so that the maximum of the common subdomain was assigned 1. If x^6 was less than 50, which was the most common case, we rescaled the RF and the CE, if it exceeded 45 we rescaled the TO. We then determined through linear interpolation the x-values under the RF and CE methods that had utilities $W^*(x^j)$, j = 1,..,5, where the x^j denote as before the elements of the standard sequence that was elicited by the TO method. These x-values were then compared with the elicited x-values under the TO.

For example, suppose an individual gave the following answers to the RF questions: $x_{0.125} = 5$, $x_{0.25} = 10$, $x_{0.5} = 18$, $x_{0.75} = 30$ and $x_{0.875} = 40$. Further, his value of x^6 is 40, so that the common subdomain is [0,40]. The utilities in the RF method are then rescaled such that the outcome 40 gets a utility of 1, i.e. by dividing all utilities by W(40) = 0.875. As a result, $W^*_{RF}(5) = 1/7$, $W^*_{RF}(10) = 2/7$, $W^*_{RF}(18) = 4/7$, $W^*_{RF}(30) = 6/7$. The x-value corresponding to a utility of 1/6 is then estimated by linearly interpolating between the two surrounding utilities of which the x-values were elicited, i.e. 1/7 and 2/7: $W^{*-1}_{RF}(1/6) = 5 + \frac{1/6 - 1/7}{2/7 - 1/7} * (10-5) = 5.83$. Outcomes above 45 in the TO method were not considered because then the rank-ordering $G > g > x^j > x^{j-1}$ was violated and the prospects under comparison were no longer rank-ordered. This was the case for fourteen participants. Of course, we could still compare the RF and the CE for these participants.

A problem when comparing the x-values is that except for the first question, the questions of the participants depend on their answers in the preceding questions. We corrected for this dependence by first computing the proportional match (PM) of each answer of the RF and CE methods, which is defined as follows:

$$PM = \frac{CE - Low}{High - Low} \tag{13}$$

where Low is the lower amount in the risky prospect for the CE method and the earliest time point in which an improvement in health occurred for the RF method. Similarly, High is the higher amount in the risky prospect for the CE method and the last time point in which an improvement in health occurred for the RF method. The outcomes Low and High vary per question. The idea of using proportional matches to correct for dependencies across questions was proposed by Miyamoto and Eraker (1988). Because we had five values for each participant, not all values were independent of each other and we had to compare them for each question separately. This resulted in the comparison of the answers to the five different answers by means of Wilcoxon signed ranks tests.

For each participant and each method we determined the shape of the utility function for life duration. Curvature of utility was investigated by computing the area under the normalized utility function. We classified a participant's utility function as concave (convex, linear) if this area was larger than (smaller than, equal to) 0.5.

To smoothen out response error, we also analyzed the data under the assumption that the utility for life duration was a power function: $W(x) = x^{\alpha}$. This family is widely used in medical decision making and generally gives a very good fit to experimental data. The power function is concave (convex, linear) if α < (>, =) 1. The power function was estimated by nonlinear least squares both for each participant separately and for the median data. All tests used below are nonparametric and two-sided. The individual power estimates provide another way to classify the participants. We classified a participant's utility function as concave (convex, linear) if his estimated power coefficient was less than 0.95 (was larger than 1.05, was between 0.95 and 1.05).

To investigate whether possible differences between the methods (e.g. more concavity under the CE method) were due to violations of expected utility, we reanalyzed the data of the CE method under prospect theory. Based on previous research, we assumed that death was the reference point of the participants (Bleichrodt et al., 2001), so that all outcomes were perceived as gains and loss aversion played no role. For probability weighting we adopted the parameter estimates of Tversky and Kahneman (1992). That is, we assumed that the probability weighting function was equal to $\frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}}$ with $\gamma = 0.61$. This implies that $w^{+}(½) = 0.42$. To compare the utilities under the RF and the CE we had to approximate the x-values of the RF method corresponding with the new CE utilities. This was accomplished in a similar way as in the comparison of the TO method with the other two methods, which was explained above.

5.7 Results

We excluded the data of three participants since they did not understand the choice task or were not willing to make risky choices concerning life duration because of religious reasons. This left the data of 67 individuals for the analysis. The consistency tests we performed revealed satisfactory test-retest reliability. The correlations between the indifference values elicited in the experiment and their repetition were high: 0.75 for the CE method, 0.74 for the RF method and 0.93 for the TO method.

Utility curvature

FIGURE 5.1. MEDIAN UTILITIES OF THE RF AND CE METHODS.

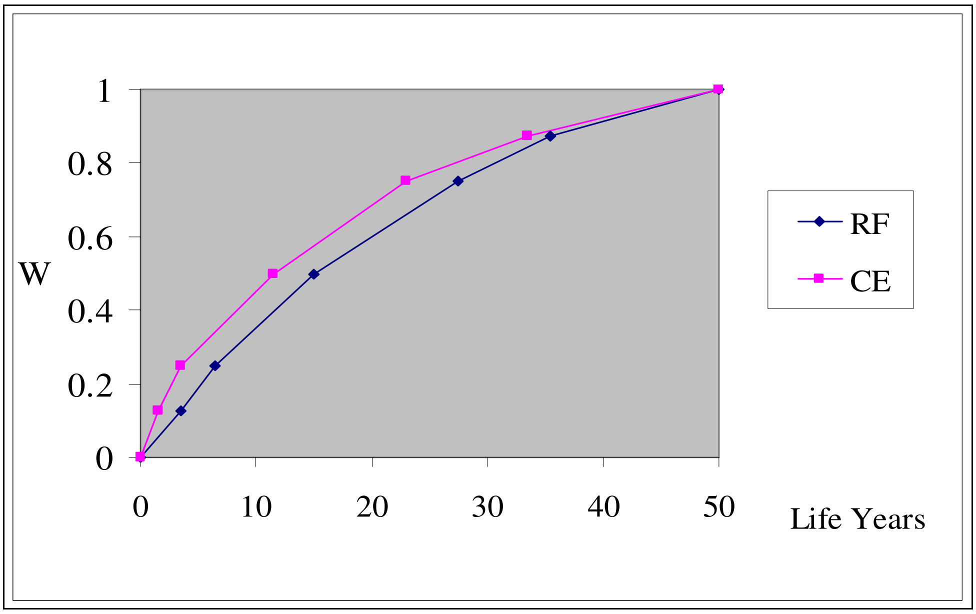


Figure 5.1 shows the utility functions for the RF and the CE based on the median data. Both utility functions are clearly concave, with somewhat more concavity for the CE. The domain of the TO method was generally smaller and we, therefore, present the median data of the TO in a separate graph (Figure 5.2). Figure 5.2 shows more variability for the TO method. It should be kept in mind though that the smaller domain of the TO is likely to produce more linearity. When we fitted a power function to the median data, the power coefficients were 0.48 for the CE, 0.62 for the RF and 0.74 for the TO. The results for the TO are comparable to those of previous studies that used the TO for measuring the utility for life duration (Bleichrodt and Pinto, 2000; Bleichrodt and Pinto, 2005).

FIGURE 5.2. MEDIAN UTILITIES OF THE TO METHOD.

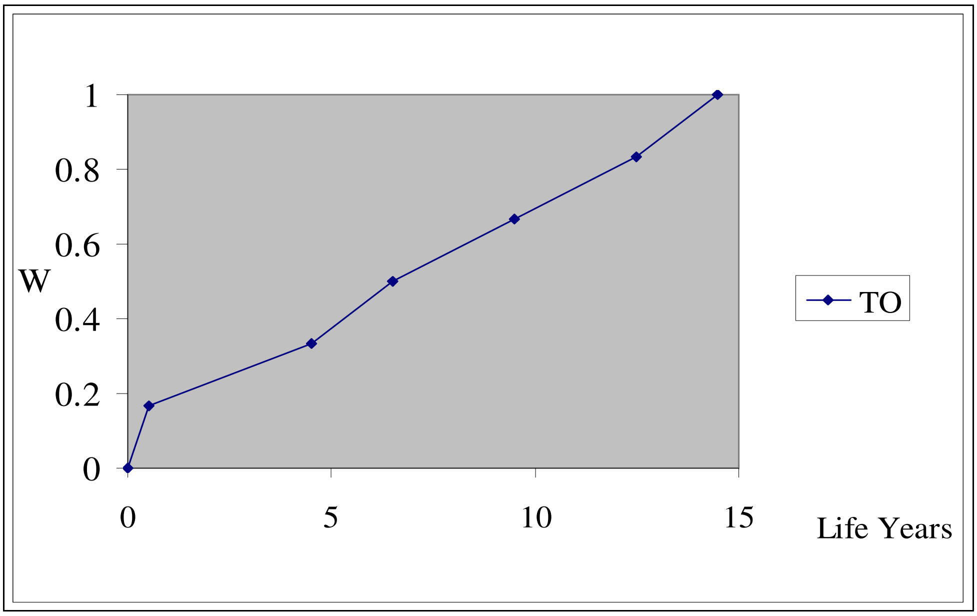


Table 5.1 shows the classification of participants according to the shape of their utility function. Concave utility is clearly the modal shape in all three methods. The difference between the number of concave and convex participants is significant in all three methods ($p < 0.001$).

TABLE 5.1. CLASSIFICATION OF PARTICIPANTS.

	RF	CE	TO
Concave	59 (88.1%)	57 (85.1%)	53 (79.1%)
Convex	8 (11.9%)	10 (14.9%)	14 (20.9%)

The medians of the individual power estimates were 0.44 for the CE, 0.62 for the RF, and 0.83 for the TO. All estimates are significantly smaller than 1, the case corresponding to linear utility ($p < 0.001$). The estimate for the CE is significantly smaller than for the RF ($p = 0.037$). Both the estimate for the CE and the estimate for the RF are significantly smaller than the estimate for the TO ($p < 0.001$). The classification of the participants based on their estimated power coefficients was very close to the nonparametric classification given in Table 5.1.

The exponential power estimates show a similar picture. The median estimate is 1.74 for the RF, 3.11 for the CE, and 0.44 for the TO. All estimates differed significantly from each other and from 0 ($p < 0.001$). An interesting result was that, for the RF and CE methods, the exponential family fitted the data significantly better than the power family (Akaike information criterion, $p < 0.001$). This fact points towards constant or even decreasing *absolute* risk aversion, instead of constant *relative* risk aversion, which corresponds with the power family. For the TO method, the power family gave a significantly better fit, but this may have been caused by the smaller domain for this method. Another interesting finding was the better fit of the RF method as compared to the CE method for both the power and the exponential family ($p < 0.001$). In addition, for the exponential family, the RF method fitted the data significantly better than the TO method ($p < 0.001$), whereas there was no significant difference for the power family ($p = 0.148$).

Direct comparison between methods

Two proportional matches differed significantly between the RF and the CE (those for $x_{0.75}$ and for $x_{0.875}$). The other three did not differ significantly ($p > 0.10$). We also compared the two methods over the whole domain by taking the differences between the proportional matches for each question and performing a Friedman test. This yielded a significant difference, indicating more concavity for the CE method than for the RF method ($p = 0.041$).

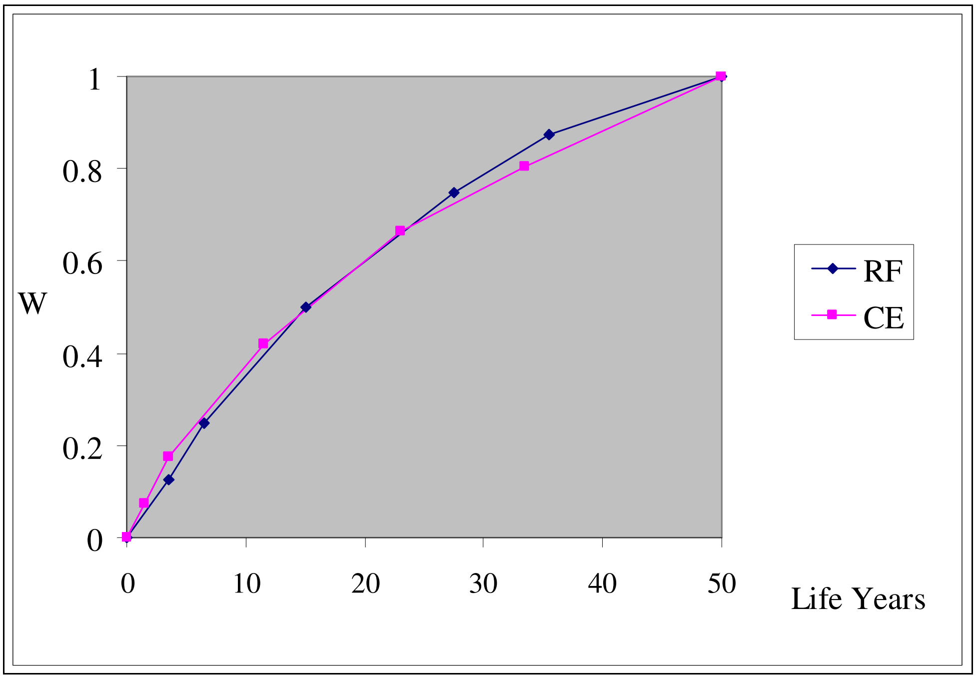
A problem with the TO method was that the value of x^6 was in general rather low and consequently the common subdomain of the methods was small. We therefore could not compare these methods over the whole domain of fifty years considered in the CE and RF methods. On their common subdomain, the difference between the RF and the TO was not significant ($p = 0.430$), but the difference between CE and TO was significant ($p < 0.001$).

Analysis in terms of prospect theory

Figure 5.3 shows the utility according to the CE when analyzed under prospect theory. The figure shows that the utility of the CE and the RF were now close. The power coefficient based on the median data was now 0.60. Remember that it was 0.62 under the RF.

No significant differences remained when we compared the x-values for CE under prospect theory with the RF ($p = 0.777$) and with the TO ($p = 0.061$). When we compared the median of the individual power estimates, no difference was observed between CE (new estimate 0.57) and RF ($p = 0.798$), but the CE estimates were still significantly lower than the TO estimates ($p < 0.02$). The classification of the participants was not markedly affected when we evaluated the CE under prospect theory: 49 participants were concave and 18 convex by the nonparametric classification and 48 concave, 15 convex and 4 linear based on the power estimates.

FIGURE 5.3. MEDIAN UTILITIES OF THE RF AND CE METHODS UNDER PROSPECT THEORY (RP=0).



Convenience

The RF method was rated as the most convenient method (mean rating 4.76), followed by the TO method (mean = 4.31). The CE was rated the least convenient (mean = 3.88). The difference in rating between the RF and CE method was significant at the 1% level (p = 0.001), while the other differences were significant at the 5% level (RF-TO, p = 0.050; TO-CE, p = 0.039). Correspondence with the participants made clear that the main reason they found the RF method easier to answer was the difficulty they had with the possibility of immediate death in the other methods. The fact that the CE method contained more questions with this possibility than the TO method probably caused the former method to be valued the least convenient.

5.8 Discussion

We observed significant differences between our new RF method, the CE and the TO. The differences between RF and CE could be due to risk aversion. The difference between CE and TO suggests that people violate expected utility and adds to the large literature on deviations from expected utility. When we evaluated the CE under prospect theory the differences between RF and CE and most of the differences between CE and TO disappeared. This suggests that differences between riskless and risky measures that have been observed in past studies were mainly due to violations of expected utility and that there is one concept of utility that can interchangeably be used in different decision contexts. This is an important finding for health economics and medical decision making where such transferability of utility is commonly used. Our finding is in line with Stalmeier and Bezembinder (1999) who previously observed that the difference between riskless and risky utility vanishes under prospect theory.

The TO produced more linearity than the other two methods. This may be due to the smaller domain of the TO method. It is well known that for small domains utility is close to linear. This caveat should be kept in mind when comparing the TO method with the other two methods.

Our analysis under prospect theory assumed that participants consider all life durations in the CE as gains. This assumption was based on previous findings by

Bleichrodt et al. (2001). The evidence of Bleichrodt et al. (2001) was obtained in a matching task. In a choice-based task van Osch et al. (2006) observed that a substantial proportion of their participants took the sure outcome as their reference point. Then the risky prospect in the CE is a mixed prospect and loss aversion affects preferences. Assuming that people take the sure outcome as their reference point, makes the utility of the CE much more linear. The median power estimate increases to 0.88 and significant differences are observed between the CE and the RF and the TO. These observations underline the need for more research into the formation of reference points in medical decision making.

Let us finally discuss the implications of our study. Our findings suggest that it is possible to measure the utility of life duration without using the problematic outcome death. This seems an important advantage as our respondents indicated that they considered our RF method simpler and more pleasant to answer than the TO and, particularly, the CE. The questions in the RF are easier to imagine for individuals, as confirmed by the participants' ratings of the methods and their expressed thoughts about them.

A second advantage of our method is that it is not prone to the biases accruing from violations of expected utility. When we consider these biases normatively irrational, it seems natural to avoid these biases wherever possible in eliciting preferences for prescriptive purposes. One possibility to accomplish this is to use some kind of correction, as we did in the reanalysis of the CE method. It seems more efficient, however, to circumvent these biases entirely by employing a method that is not influenced by them, like the RF method.

Appendix 5A. Health state descriptions (translated from Dutch)

Card 1 – Regular Back Pain

You have regular back pain. This has the following consequences for your functioning in daily life:

- You have no problems in walking about.
- You have no problems to wash or dress yourself.
- You have *some* problems with your usual activities.
- You have *moderate* pain or other discomfort.
- You are not anxious or depressed.

Card 2 – Full Health

You have no complaints and are in perfect health. This has the following consequences for your functioning in daily life:

- You have no problems in walking about.
- You have no problems to wash or dress yourself.
- You have no problems with your usual activities.
- You have no pain or other discomfort.
- You are not anxious or depressed.

Appendix 5B. Display screen

Appendix 5C. Explanation midpoint technique

The midpoint technique is illustrated in Table 5A. There we use the stimuli of the first iteration process of the RF method, i.e. the elicitation of $x_{0.5}$ and show the answers of an imaginary participant. The option he chooses is indicated in italics. We always started with a value that was in the middle of the two extreme values and then adjusted this starting value upwards or downwards depending on the option chosen. The size of the change was always half the size of the change in the previous question, with the restriction that the numbers were rounded to integers. The method resulted in an interval within which the indifference values should lie. The midpoint of this interval was taken as the indifference value. For the CE and the TO method, this technique was applied in a similar way.

TABLE 5A. ILLUSTRATION OF THE MIDPOINT TECHNIQUE.

Iteration	Offered choices in the elicitation of $x_{0.5}$
1	**<u>0 to 25</u>** or 25 to 50
2	0 to 13 or **<u>13 to 50</u>**
3	**<u>0 to 19</u>** or 19 to 50
4	0 to 16 or **<u>16 to 50</u>**
5	0 to 18 or **<u>18 to 50</u>**
Indifference value	**<u>18.5</u>**

Appendix 5D. Convenience questions (version RF-TO CE, translated from Dutch)

Indicate, on a scale from 1 up to and including 7, to what degree you agree to each of the beneath propositions. The number 1 means that you fully DISAGREE and the number 7 means that you fully AGREE.

1. The first part of the experiment, where you had to choose between treatments that relieve your back pain during different periods, was easy to answer.

2. The second part of the experiment, where you had to choose between two risky options, 1 with a smaller difference between the two possible outcomes, and 1 with a larger difference between the two outcomes, was easy to answer.

3. The third part of the experiment, where you had to choose between a sure period that you live and a treatment with a chance for a longer period to live and a chance for a smaller period to live, was easy to answer.

6 The correction of TTO scores for utility curvature using a risk-free utility elicitation method[1]

Summary

The TTO method has become the most frequently used method to measure health state utilities. However, it is still influenced by some important biases. One of these is the role of nonlinear utility of life duration, commonly believed to be caused by the discounting of future health and diminishing marginal utility of additional life years. Ignoring utility curvature in the process of valuing health states can result in a downward bias in health valuations. Moreover, if respondents already discount future life years in health state valuations, applying a standard discount rate to TTO estimates in economic evaluations, will result in double discounting. This chapter describes and employs a new method to correct TTO estimates for utility curvature. Unlike most previous attempts, the method we use is risk-free. It is robust to several biases that occur under methods that

---

[1] This chapter is based on Attema and Brouwer (2007a).

incorporate risk. Our results show a significant degree of utility curvature in TTO estimates. The risk-free method seems to be useful to correct TTO estimates for this influence and leads to significantly higher health state valuations.

6.1 Introduction

The time tradeoff (TTO) method is a popular way of eliciting preferences for health states (e.g. Dolan, 2000). As a consequence, several quality of life tariffs used in cost-effectiveness analyses are based on TTO measurements (e.g. Dolan, 1997; Lamers et al., 2006). In a TTO individuals need to make a tradeoff between quality of life and duration of life. A typical TTO exercise will involve a tradeoff between living in some imperfect health state β for 10 years and living in perfect health for a period less than 10 years. The amount of time people are willing to sacrifice in order to restore perfect health then indicates the value of the health state β and can subsequently be used to calculate the quality adjusted life year (QALY) score of that health state.

Despite its popularity, the traditional TTO method has been shown to be prone to several potential biases. The method makes strong assumptions such as linear utility of life duration, no loss aversion and no scale compatibility, which are hard to maintain (e.g. Nord, 1992; Bleichrodt, 2002). Consequently, the QALY scores elicited by the conventional TTO procedure are biased. Loss aversion and scale compatibility cause an upward bias in QALY scores (Bleichrodt, 2002). Moreover, utility of life duration is often found to be nonlinear, which mainly relates to two aspects: (i) diminishing marginal utility of additional lifetime and (ii) discounting. Both are problematic in the context of a TTO, as this method does not take into account utility curvature, leading to a downward bias in QALY scores (Bleichrodt, 2002).

Diminishing marginal utility refers to the fact that the utility derived from an additional unit of some good declines with the quantity of that good that an individual already possesses. This implies that the utility increase from having the projected tenth year in the TTO is lower than that from the first year. Another important issue in the TTO method is that of discounting. A typical respondent having to trade off future life years in order to restore full health is likely to discount future life years (e.g. Stiggelbout et al., 1994; Stalmeier et al., 1996; Wakker and Deneffe, 1996; Martin et al, 2000; Bleichrodt and Pinto, 2005; van der Pol and Roux, 2005). Discounting implies, as Bohm-Bawerk already put it, that: '*To goods that are destined to meet the wants of the future, we ascribe a value which is really less than the true intensity of their future marginal utility*' (as quoted in Olsen, 1993). Both mechanisms cause a lower value to be attached to the future life years that are traded off in a TTO exercise. This immediately indicates the problem that this chapter addresses. Simply using the number of future life years that individuals are willing to trade off in calculating QALY scores misrepresents the utility attached to a current imperfect health state. In order to have a better estimate of the true valuation of a health state, a correction for utility curvature is required, therefore. This is especially true for discounting given the way that resulting health state valuations are normally used in economic evaluations, i.e., they are discounted to calculate a net present value of QALYs (e.g. Gravelle et al., 2007). If uncorrected TTO values are used to calculate

QALYs and these are subsequently discounted using some discount rate for health effects, this would amount to double discounting and an underestimation of the utility derived from some health state (MacKeigan et al., 2003).

This chapter focuses on the role of nonlinear utility of life duration in TTO exercises and describes a new method to correct for utility curvature. This involves a recently proposed riskless method that does not need to make specific parametric assumptions about the utility function or discounting behavior (Attema et al., 2007).

The structure of the chapter is as follows. First, we introduce the theory underlying our study in Section 6.2. We also discuss related literature concerning adjusted TTO scores there. In Section 6.3 we explain the method used to elicit utility for life duration and the way to use this information to correct raw TTO scores. The experimental details are put forward in Section 6.4, followed by a presentation of the results in Section 6.5. Finally, Section 6.6 discusses the results and concludes.

6.2 Theory and related literature

A common way to describe preferences over lifetime utility is to represent them by the following multiplicative utility function over life duration and health quality:

$$V = \sum_{t=j}^{T} \delta_t u(h_t) \tag{1}$$

with $u(h_t)$ a utility function that represents the individual's preferences over health states at each time point t, δ_t denoting the corresponding weight attached to the utility at this point, j the starting period, and T the complete time frame. An axiomatic derivation for this model, which we will call the *generalized QALY model*, was given by Bleichrodt and Gafni (1996).

The conventional TTO method is embedded, however, in a special case of the generalized QALY model, which we will call the *linear QALY model*. In this linear QALY model, it is assumed that equal weight is attached to all utilities regardless of their timing, so that $\delta_t = 1$ for each t in Equation (1). This is a restrictive assumption, however, since it implies no utility curvature for life duration. Given this assumption, the conventional TTO method measures the utility of a health state β by asking the respondent to give some period in full health, followed by death, which makes him indifferent to a stated period in health state β, also followed by death. As a result, the elicited indifference can be represented by the following equation:

$$n_\beta * u(\beta) = n_{FH} * u(FH) + (n_\beta - n_{FH}) * u(D) \tag{2}$$

with *FH* representing full health and *D* death, n_β is the stated period in health state β, and n_{FH} is the elicited period in full health. If the utility function over health is normalized so that $u(FH) = 1$ and $u(D) = 0$ we get the following simple expression for $u(\beta)$:

$$u(\beta) = \frac{n_{FH}}{n_{\beta}} \tag{3}$$

However, this expression only holds if the utility over life duration is indeed linear and there is no reallocation of lifetime consumption due to the smaller duration in full health. The latter concern will probably not cause a large bias (Dolan and Jones-Lee, 1997), but as indicated in the introduction, the assumption of linear utility over life duration does not seem realistic. If we take into account the existence of nonlinear utility of life duration, the estimate obtained by Equation (3) will clearly be biased and the resulting bias can be substantial (e.g. MacKeigan et al., 2003). Moreover, given the normal practice to discount health effects in economic evaluation, using these uncorrected TTO scores will result in double discounting of health effects.

It seems necessary, therefore, to avoid the restrictive assumptions of the linear QALY model. Simply starting from the generalized QALY model, and not imposing any restrictions on the weight respondents might attach to utilities over time, δ_t, the indifference implied by the TTO method (using the same notation as before) would give the following equation:

$$\sum_{t=0}^{n_{\beta}} \delta_t u(\beta) = \sum_{t=0}^{n_{FH}} \delta_t \tag{4}$$

Measuring the value of *u(β)* therefore requires to have estimates of $\sum_{t=0}^{n_\beta} \delta_t$ and $\sum_{t=0}^{n_{FH}} \delta_t$ first. The purpose of this chapter is to estimate corrected TTO values using the risk-free (RF) method proposed by Attema et al. (2007), which measures utility for life duration without making assumptions about δ_t. Therefore the method should elicit TTO values that are not distorted by utility curvature.

There have been some previous attempts to correct TTO scores for the utility of life duration (Pliskin et al., 1980; Stiggelbout et al., 1994; Stalmeier et al., 1996; Martin et al., 2000; van Osch et al., 2004; van der Pol and Roux, 2005). However, most of these studies considered a risky situation in which they elicited certainty equivalents (CEs) to get a measure of the utility of life duration. That is, respondents were asked one or more questions where they had to indicate a sure amount of remaining life years that made them indifferent to a 50-50 gamble consisting of one higher and one lower amount of remaining life years.

Pliskin et al. (1980) asked one CE question where the gamble consisted of 5 and 15 years in a constant mild health state. Their estimates were close to linearity. Pliskin et al. (1980) did not explicitly test for differences between unadjusted and adjusted TTO scores. Stiggelbout et al. (1994) elicited three CEs, with immediate death as the lowest possible outcome and 10 years as the highest possible outcome. A power function was used to fit these data and the estimated power was applied to correct the raw TTO scores. Their median power estimate was 0.73 and the adjusted TTO scores were significantly higher than the raw scores. In the study of Stalmeier et al. (1996), respondents performed seven life-year gambles in good health, with death within a month as the lowest possible outcome. They fitted the utility function with a power, logarithmic, exponential or logistic function and reported mainly concavity of the utility function for life duration and a substantial difference between unadjusted and adjusted TTO scores. Martin et al. (2000) asked three CE questions in a sample of cardiovascular disease patients, one gamble between 1 and 10 years, one between 5 and 15 years, and one between 10 and 20 years. They therefore did not include immediate death. The results were fitted with a power and an exponential function. There was substantial concavity and the exponential function gave a better fit than the power function, which is evidence in favor of a constant absolute risk attitude instead of a constant relative risk attitude. The exponential function was subsequently used to correct the raw TTO scores upwards. A difference varying from 0.02 to 0.06 was found, depending on the durations used. Finally, the procedure used by van Osch et al. (2004) was the same as that of Stalmeier et al. (1996), except that death within a month was replaced by death within a week. They used the answers to estimate a power utility function. They found utility for life years to be nearly linear at the aggregate level. As a result, their corrected TTO scores were only slightly higher than the uncorrected ones.

These studies show that utility curvature can be important and that a correction of TTO scores for curvature can result in significantly higher health state valuations, although this is not confirmed in all studies. The latter will probably be related to differences in methods and populations between the studies. An important feature of all the above studies is that they use a method entailing risk to derive utility curvature. Such an approach, however, may be considered to be less compatible with the TTO method, which does not entail risk. Moreover, using risky situations in deriving utilities of life duration can lead to a distortion of results due to probability weighting, which may bias the utilities obtained by the certainty equivalence method (Abdellaoui et al., 2007). Finally, several authors have pointed out that the inclusion of immediate death leads to extreme risk aversion and, hence, to strong concavity of utility (e.g. Tversky and Kahneman, 1986; Shin et al., 1997; Stiggelbout and de Haes, 2001; Bleichrodt et al., 2003). It therefore seems worthwhile to find risk-free methods to obtain estimates of utility curvature for life duration. Such methods seem more compatible with the risk-free nature of the TTO method, and in addition avoid probability weighting biases and the problematic inclusion of immediate death as an outcome.

One earlier attempt to derive corrected TTO values using a risk-free method has been reported to our knowledge. Van der Pol and Roux (2005) elicited time preference for an increase in body weight. However, in contrast to the present study, they asked only one question to measure time preference and therefore had to assume specific parametric discounting models to correct their TTO scores. They used an open-ended question to elicit time preference, where respondents had to specify the number of years with their weight being 20% higher starting in 45 years time that was just as bad as their weight being 20% higher for 5 years starting in 15 years. This answer was used to estimate the parameter of the constant discounting model and of a specific hyperbolic discounting model. Subsequently, the raw TTO scores were adjusted upwards with these estimates. Van der Pol and Roux (2005) report a significant difference between raw and adjusted scores, with their mean estimated adjustment factor lying around 0.03 (4.4%).

In this paper we use a recently developed method to derive utility curvature (Attema et al., 2007). In comparison to the method used by van der Pol and Roux (2005) to derive utility curvature, which may be perceived to be cognitively demanding for lay respondents, our method is relatively easy to use. The comprehensiveness of the RF method was also confirmed in a questionnaire of Attema et al. (2007), which indicated that respondents found this method significantly easier to answer than the CE method. Moreover, the method relates well to TTO tradeoffs, since we make use of a choice-based procedure, requiring respondents to make tradeoffs between health profiles. Given these tradeoffs, the procedure is better embedded in economic theory than choiceless procedures (Dolan, 2000). Moreover, the method is developed in such a way that more measurements of the utility function can be obtained, allowing an accurate correction of the TTO results. Finally, our results are less susceptible to the validity of a particular parametric shape of the utility function for life duration, as will be further highlighted below.

6.3 Method

The full elicitation method consists of two distinct parts. First of all, we measure the degree of utility curvature. Then, in the second phase we perform a conventional time tradeoff. The results from the first phase are used to correct the responses in the second phase. Given that the conventional TTO method has already been discussed above, we focus here on the first phase and the correction of the answers in the second phase.

The first phase uses the method of Attema et al. (2007). We briefly summarize their method here. The respondents' task is to compare two different health profiles. The two health states are β and γ, with $\gamma \succ \beta$. In the first health profile, *A*, the respondent gets an immediate improvement in health from β to γ, which lasts until time point *m*, after which the respondent returns to health state β until point *T*. In the second health profile, *B*, he will remain in this basic health state until time point *m* and then gets the health improvement, the latter lasting until point *T*. After point *T* the two options yield the same health state. Now we vary *m* until the respondent is indifferent between these two options. We then obtain the following equation:

$$\sum_{t=0}^{m} \delta_t H(\gamma) + \sum_{t=m}^{T} \delta_t H(\beta) = \sum_{t=0}^{m} \delta_t H(\beta) + \sum_{t=m}^{T} \delta_t H(\gamma) \tag{5}$$

As shown by Attema et al. (2007), this implies[2]:

$$\sum_{t=0}^{m} \delta_t = \frac{1}{2} \tag{6}$$

---

[2] We assume that it does not matter whether the respondent is presented with an increasing health profile over time first or rather with a decreasing health profile first in this exercise. In other words, replacing the left-hand side of Equation (5) by the right-hand side and vice versa will not lead to differences in the estimation of point *m*. Our results were tested for this and confirm this hypothesis.

We proceed by using this estimate of *m* at the place of *T* and posing a similar question in order to infer the respondents' value of *k* such that $\sum_{t=0}^{k}\delta_t = \frac{1}{4}$. That is, in profile *A*, the immediate improvement in health from β to γ now lasts until time point *k*, after which the respondent returns to health state β until point *m*. In profile *B* he will remain in this basic health state until point *k* and then gets the health improvement, which now lasts until point *m*. In a similar way, when we replace the 0 of Equation 5 by *m*, we can infer the value of *q* such that $\sum_{t=0}^{q}\delta_t = \frac{3}{4}$. Continuing in this manner, we can get more detailed information about the shape of δ_t. We can do this without making particular assumptions about the functional form of δ_t. Once we have elicited an estimate of δ_t using this method, we can use it to correct TTO estimates for utility of life duration.

To illustrate the correction of the TTO scores, consider the following situation. Let us assume that the purpose of the exercise is to estimate the utility of health state β. We may then ask the respondent, as this is normally done in TTO exercises, to specify the number of periods, n_{FH}, in full health that makes him indifferent to *m* years in health state β. Making use of (6) this gives:

$$\frac{1}{2}u(\beta) = \sum_{t=0}^{n_{FH}}\delta_t \tag{7}$$

Suppose that the elicited value of n_{FH} lies somewhere between the points *k* and *m* of the integral. By linear interpolation we can then estimate the value of $\sum_{t=0}^{n_{FH}}\delta_t$ by computing $\sum_{t=0}^{k}\delta_t + \frac{n_{FH}-k}{m-k}\sum_{t=k}^{m}\delta_t = \frac{1}{4}\left(1+\frac{n_{FH}-k}{m-k}\right)$. If, for example, we have $m = 10$, $k = 4$, and $n_{FH} = 8$, we get $\sum_{t=0}^{8}\delta_t = \frac{1}{4}\left(1+\frac{2}{3}\right) = \frac{5}{12}$. Using Equation (7), it is easy to derive that $u(\beta) = \frac{5}{6}$.

This also makes clear that using uncorrected TTO scores results in an underestimation of the utility attached to health state β if utility of life duration is concave, since the conventional TTO method would have given a value of 8/10, which is smaller than the value of 5/6 we found above. Similarly, convex utility of life duration would result in an overestimation of the utility of β.

6.4 Experiment

In this section, we present an experiment, using the method described above.

Participants

Seventy participants were recruited and were paid a fixed amount of €12.50 to join the experiment. The participants were students from different faculties of the Erasmus University Rotterdam. Before the actual experiment, we tested the design in several pilot sessions using other students and university staff as participants.

Procedure

The experiment was administered in sessions of at most two persons each using a computer program. During each session there was an experimenter in the laboratory to give instructions and clarify possible opacities. The questions considered in this study were part of a larger experiment, which lasted no longer than 45 minutes. Further details about the utility elicitation part of the experiment can be found in Chapter 5.

Stimuli of the RF method

Health state β in the RF method was specified as regular back pain. We selected this health state because it is a common illness and the participants were likely to know people suffering from it. We described the health state using the domains contained in the EuroQol 5D questionnaire. We therefore indicated what regular back pain meant for daily functioning in terms of five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The descriptions were printed on cards and handed to the participants (see Appendix 5A). Health state γ was specified as full health. It was made clear to the participants that this health state meant they were able to function perfectly on all five dimensions, irrespective of their age.

T was set equal to 50 years, because this was a plausible amount for our sample of students. In the first question we determined $x_{0.5}$ such that $\sum_{t=0}^{x_{0.5}} \delta_t = \frac{1}{2} \sum_{t=0}^{50} \delta_t = \frac{1}{2}$ by means of a choice-based procedure. We did this by eliciting the point where the participant was indifferent between a health profile starting with full health until time $x_{0.5}$, followed by back pain for the remainder of the 50 years and a health profile starting with back pain until time $x_{0.5}$, followed by full health for the remainder of the 50 years. We told the participants that after this period the two options would be followed by the same health state without specifying it further. After having elicited the point $x_{0.5}$, we subsequently used this point in the next question. In this way we elicited $x_{0.125}$, $x_{0.25}$, $x_{0.75}$ and $x_{0.875}$.

Stimuli of the TTO method

The conventional TTO procedure to elicit the value of some health state β is to let the participants state the number of years n_{FH} in full health that they consider equivalent to a specified duration n_β in health state β, i.e. such that $n_{FH} \sim n_\beta$. The unadjusted TTO value is then given by n_{FH}/n_β. Our experiment entailed two TTO questions of this kind. The health state (β) that we valued was specified as regular back pain again. We fixed the duration n_β at 14 (BP14) in one question and 27 years (BP27) in another one. The order of these questions was randomized.

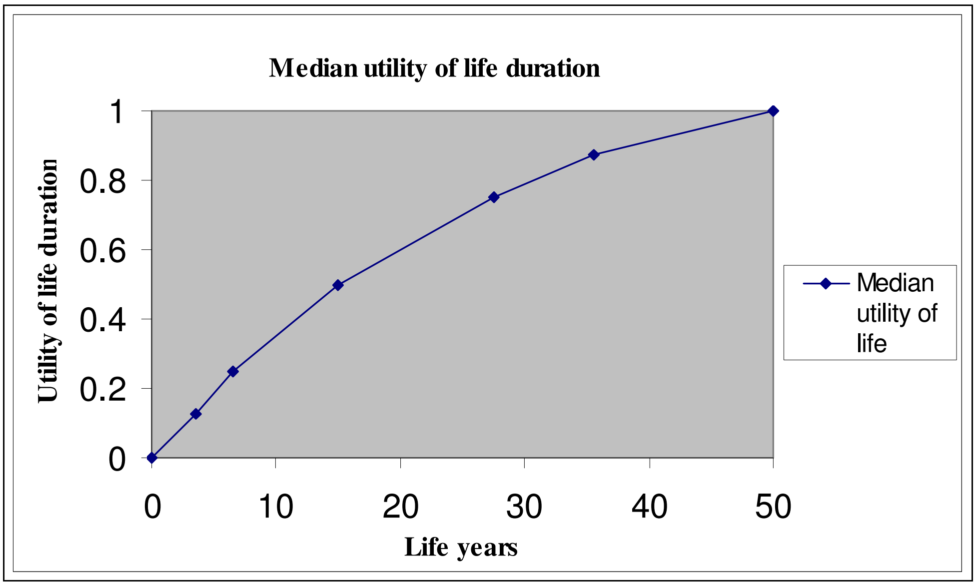
6.5 Results

Fifty-six participants were included in the analyses (mean age 21.8 years, range 18-37 years, 20 (35.7%) females). The other 14 participants were eliminated from the sample because they had at least one answer not corresponding to their reasoning. Two of these eliminated participants did not fully understand the utility elicitation phase, whereas 12 participants were removed because they had difficulties completing the TTO phase.[3]

Utility for life duration

The first phase of the experiment was aimed at deriving the utility for life duration. In Figure 6.1 we show the median data of the elicited utility for life duration functions.

FIGURE 6.1. MEDIAN UTILITY OF LIFE DURATION ESTIMATES.



[3] For example, for one participant indifference between 10 years in full health and 5 years with back pain was elicited in the TTO part, although this participant clearly preferred the former alternative to the latter when she was asked to choose between them.

The figure shows a clearly concave pattern, indicating that years farther in the future receive less weight compared to years nearer in time. This confirms that respondents indeed may discount future life years and may (expect to) experience diminishing marginal utility of additional life years. A convex utility function on the other hand would reflect more attention to utility in the future, or negative discounting. In order to give a more detailed description of the degree of concavity and convexity, we classified participants as concave or convex depending on their five answers to the utility elicitation questions. This was done by computing the differences between two successive elements of the elicited utility points and dividing these by their respective utility increase (this division was necessary since the utility difference between two successive elicited values could be either 0.125 or 0.25):

$$\Delta_j = \frac{x_j - x_{j-1}}{\sum_{t=x_{j-1}}^{x_j} \delta_t}, j = 1,\ldots,6 \tag{8}$$

Then we computed:

$$\partial_j = \Delta_j - \Delta_{j-1}, j = 2,\ldots,6 \tag{9}$$

i.e. how much successive outcome intervals increase or decrease per utility unit. For each participant we observed five values of ∂_j. A positive value of ∂_j corresponds to a concave part of the utility function. It means that an individual needs a larger increase in life duration to obtain a given increase in utility at higher amounts than at lower amounts. Likewise, a negative value of ∂_j corresponds to a convex part of the utility function and a value of 0 corresponds to linear utility. We classified a participant as having linear (concave, convex) utility if he had at least three out of five linear (concave, convex) parts. This criterion was used to account for response error. If none of the three parts (linear, concave or convex) occurred more than twice, the participant was not classified. Table 6.1 shows the results. It is clear that most participants showed positive discounting of future life years, which causes a downward bias in raw TTO scores, i.e. an underestimation of utility.

TABLE 6.1. UTILITY CLASSIFICATION OF PARTICIPANTS.

	Number
Concave	48 (85.7%)
Convex	7 (12.5%)
Linear	1 (1.8%)
Unclassified	0 (0%)
Total	56

To give more insight into the degree of concavity, we estimated a power ($u = x^r$, corresponding to a constant relative risk attitude over life years) and an exponential ($u = \frac{1-e^{-cx}}{1-e^{-c}}$, corresponding to a constant absolute risk attitude) utility function for the data of each participant. These parametric families were estimated by nonlinear least squares. The exponential family fitted the data significantly better than the power family (Akaike information criterion, $p < 0.001$) and therefore we only report the results of the exponential function here. Actually, the data seemed to indicate an increasing absolute risk aversion instead of a constant absolute risk attitude for life duration, so that the exponential function may still not give the best fit. Note, however, that our correction procedure for TTO scores does not require the estimation of this function, as it depends only on the observed data points and the assumption of linearity in between two points.[4] The median estimate of the exponential utility function is 1.74, which is significantly higher than 1 ($p < 0.001$) and therefore indicates concavity.

What is clear, therefore, is that there was a clear degree of concavity in utility of life duration in this sample, which, without correction, on average leads to underestimation of the utility attached to the health state under study (abstracting from other biases).

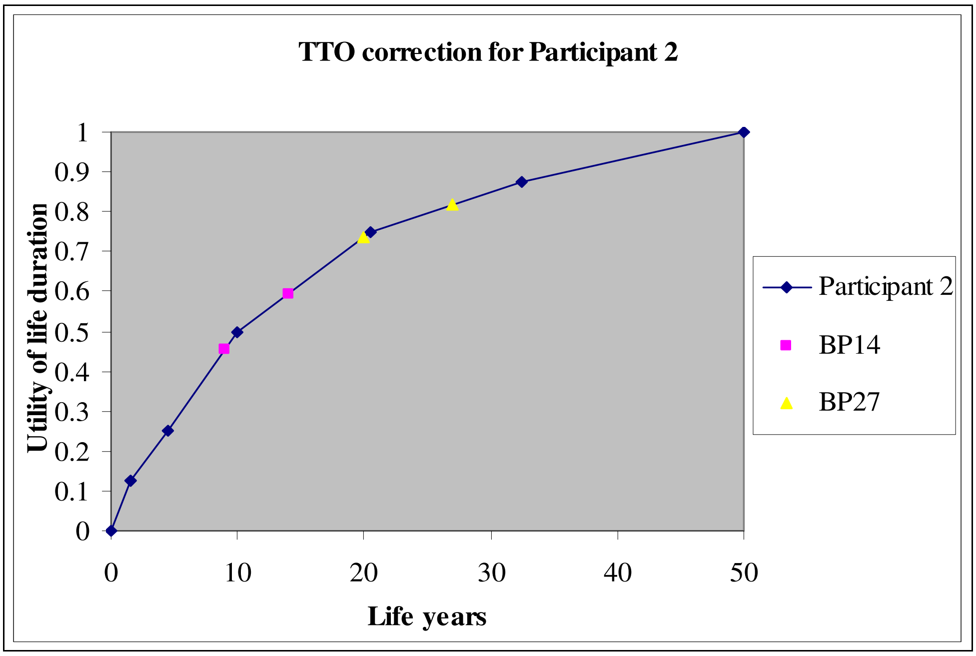
TTO scores

The data of a representative participant of the experiment may clarify the way we corrected TTO scores of each individual for their utility of life duration curvature. Let us take a look at Participant 2. This participant had the following utility points: $x_{0.125} = 1.5$, $x_{0.25} = 4.5$, $x_{0.5} = 10$, $x_{0.75} = 20.5$ and $x_{0.875} = 32.5$. As can be seen in Figure 6.2, this participant clearly had a concave utility function. The finding that $x_{0.5} = 10$, for example, means that this participant values the next 10

[4] We did not attempt to estimate functions that take into account increasing absolute risk aversion, since it is not pivotal for the present paper and rather laborious (e.g. Martin et al., 2000).

years equally as the 40 years thereafter, whereas 10 years is only one fourth of 40 years. That is, the weight attached to years in the short term is higher than that attached to later years.

FIGURE 6.2. EXAMPLE OF THE UTILITY CORRECTION OF TTO SCORES.



In the TTO part, Participant 2 gave the following answers: BP14 = 9 and BP27 = 20. These numbers are also included in Figure 6.2. Obviously, the raw TTO scores for this participant are 9/14 = 0.6429 for BP14 and 20/27 = 0.7407 for BP27.

For the corrected TTO scores, the computation is somewhat less straightforward. There we have to compute the ratio of the weighted answers, instead of computing the ratio of just the answers. In order to do so, for BP14 we first compute the participant's utilities of 9 and 14 life years. The utility data show that the number *9* lies in between $x_{0.25}$ and $x_{0.5}$. This implies that in order to get an estimate of the utility of the coming 9 years, $\sum_{t=0}^{9}\delta_t$, we have to interpolate between these utility points: $\sum_{t=0}^{9}\delta_t$ = 0.25 + (9-4.5)/(10-4.5)*(0.5-0.25) = 0.4545. Similarly we obtain: $\sum_{t=0}^{14}\delta_t$ = 0.5 + (14-10)/(20.5-10)*(0.75-0.5) = 0.5952.

Using these figures, we can now derive the corrected TTO score for Participant 2 for the BP14 case. This score is calculated as 0.4545/0.5952 = 0.7636, which, as expected, is higher than his raw TTO score of 0.6429. In a similar fashion we obtain a corrected TTO score for BP27 for Participant 2 of 0.9026, which again is higher than the raw score of 0.7407.

In Tables 6.2 and 6.3 we present summary statistics concerning the raw and corrected TTO scores for the full sample. In order to test for the sensitivity to the linear interpolation, we also report the corrected TTO scores using the estimated exponential utility functions as correctors. As expected, the corrected scores are higher than the raw scores ($p < 0.01$ for both questions), due to diminishing marginal utility. The difference between the raw and adjusted values is 0.04 (5.8%) for BP14 and 0.05 (6.7%) for BP27. Besides, the nonparametriccaly and exponentially corrected scores are similar ($p > 0.1763$), so that our linear interpolation does not seem to cause many problems.

TABLE 6.2. RAW AND ADJUSTED TTO SCORES (BP14).

BP14	Raw scores	Corrected scores (nonparametric)	Corrected scores (exponential)
Mean	0.7092	0.7501	0.7471
Standard error	0.0247	0.0249	0.0245
Interquartile range	0.6429-0.8571	0.6911-0.8797	0.7033-0.8767
Range	0.0714-1	0.0935-1	0.0944-1
Average number of life years traded	4.07	-	-
Difference with raw scores	-	0.0409 (5.8%)	0.0379 (5.3%)

TABLE 6.3. RAW AND ADJUSTED TTO SCORES (BP27).

BP27	Raw scores	Corrected scores (nonparametric)	Corrected scores (exponential)
Mean	0.7509	0.8015	0.8092
Standard error	0.0249	0.0254	0.0241
Interquartile range	0.7315-0.8889	0.6915-0.9353	0.7267-0.9403
Range	0.0741-1	0.1321-1	0.1229-1
Average number of life years traded	6.73	-	-
Difference with raw scores	-	0.0506 (6.7%)	0.0583 (7.8%)

6.6 Discussion

This study has reported the use of a recently developed method to derive the degree of concavity in utility of life duration and applied it to correct TTO scores. Our results provide further evidence that respondents indeed do not have a linear utility function for life duration, causing the conventional TTO method to yield downward biased estimates. The risk-free method we employed to correct the raw TTO scores performs well in our sample. Its form and stimuli are similar to that of the TTO and, therefore, the method seems promising as a means of correcting TTO scores for utility curvature. Note that although a considerable fraction of respondents (20%) was removed from the analysis, this was mainly caused by difficulties in understanding the conventional TTO procedure and to a much smaller extent by problems with the utility of life duration elicitation procedure. Only 2.9% of the sample had to be removed because of difficulties with the RF method.

Some aspects of the method deserve mentioning. First, a potential drawback of this nonparametric risk-free method is that one has to assume linear utility between two measured utility points. However, if this is considered a problem, one can reduce the reliance on interpolation by obtaining more measurements and therefore reducing the space between observed points. This will of course require a more elaborate measurement procedure. Besides, by comparing the nonparametric estimates to parametric estimates, one can look at the potential consequences of the linearity assumptions. In our study, this did not cause large differences. A second point is that, in our study, we have applied one particular health state (back pain) in both the TTO exercise and the risk-free method eliciting utility curvature. This was done deliberately, since we cannot exclude the possibility that different utility of life duration functions exist for different health states. In other words, the degree of utility curvature may depend on the health state used in the health profiles. It is, for instance, well known that losses and gains are discounted differently and that the size of gains and losses may be of importance as well (e.g. Thaler, 1981; Benzion et al., 1989; Abdellaoui et al., 2006). This implies that there might not be a single utility function over life duration applicable to all health states. If it is confirmed that the degree of concavity depends on the health state presented, it seems worthwhile to combine the here described method and TTO exercises in such a way that for the health state under study the relevant utility of life duration function is elicited. On the other hand, there is considerable support for utility independence, so that this problem may not be too worrisome (e.g. Bleichrodt and Pinto, 2005). It would be worthwhile investigating this aspect further in future research.

In this study we found a significant and substantial difference between raw TTO scores and TTO scores that were corrected for utility curvature. The results are similar to those of Stiggelbout et al. (1994), Stalmeier et al. (1996), Martin et al. (2000) and van der Pol and Roux (2005), and therefore expand the evidence of the presence of a downward bias in the ordinary TTO method that might compensate the upward biases due to loss aversion and scale compatibility. In addition, in Chapter 5, for the participants of the present study the utility for life duration was also elicited by means of the CE method. This method yielded more concavity than the RF method, so a correction of the TTO scores using those utility estimates would probably cause an even higher upward adjustment. The only exception we could find concerns the results of van Osch et al. (2004), which suggest that the effect of correcting TTO for utility curvature is only minor. A majority of the empirical evidence thus suggests that discounting raw TTO scores will lead to an underestimation of true preferences for health states.

Of course, the bias caused by utility curvature may be partly balanced or even offset by upward biases caused by loss aversion and scale compatibility. One could argue, therefore, that correcting for the bias of utility curvature is unnecessary or even undesirable. However, we would argue that a more precise elicitation of health state valuations requires avoiding biases or correcting for them rather than relying on these biases to somehow equal out. The method proposed here offers a clear way of correcting for an important bias in TTO. Moreover, as pointed out in Chapter 5, the RF method makes it possible to reliably measure the utility of life duration without needing the problematic outcome immediate death. That chapter also reported evidence that respondents find this method simpler and more pleasant to answer than alternative risky elicitation methods, like the certainty equivalence method. Further, the method avoids biases like probability weighting and does not make parametric assumptions. Finally, it seems more natural to use in a TTO context as the TTO method considers a riskless situation.

We believe, therefore, that the method presented here provides an interesting and tractable way of correcting TTO answers for the downward pressure of utility curvature. Although TTO may be prone to other biases as well, rather than relying on these biases to result in some kind of optimal balance, it seems worthwhile to attempt to find solutions. This chapter serves to contribute in that quest.

7 Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO measurement[1]

Summary

The TTO method has often been used to value health states, but it is susceptible to several biases and methodological difficulties. One of these is a violation of procedural invariance, which means that the way a TTO question is framed can have a substantial effect on the elicited value of a health state. There are four important sources of discrepancy of the two procedures: loss aversion, maximum endurable time, scale compatibility and utility curvature (mostly due to discounting). In this chapter we present the results of a new test of procedural invariance in which we avoided or corrected for two of these sources (discounting and maximum endurable time). Our results indicate that while correcting for discounting does diminish the difference between the two TTO procedures, a large and significant violation of procedural invariance remains. Loss aversion is probably the main determinant of the remainder of this difference.

---

[1] This chapter is based on Attema and Brouwer (2007b).

7.1 Introduction

The time tradeoff (TTO) method is a popular method to value health states and many quality of life tariffs used in cost-effectiveness analyses are based on TTO valuations (e.g. Dolan, 1997; Dolan, 2000; Lamers et al., 2006). In spite of their popularity, however, TTO scores are influenced by several biases, like loss aversion and scale compatibility (e.g. Bleichrodt, 2002). Moreover, there are questions regarding respondents who are unwilling to trade off length of life to gain quality of life (e.g. McNeil et al., 1981; Stiggelbout et al., 1995) and regarding the influence of discounting on TTO scores (e.g. Dolan and Jones-Lee, 1997; Bleichrodt, 2002; van der Pol and Roux, 2005; Attema and Brouwer, 2007a). A fairly new research area in this context is that of procedural invariance of the TTO method (e.g. Spencer, 2003). The *conventional TTO procedure* fixes the number of years in some health state β and subsequently elicits the number of years in full health that would make a respondent indifferent between the two options. However, the value of health state β might as well be measured by an *alternative TTO procedure* in which the number of years in full health is fixed and subsequently the number of years in health state β that would make a respondent indifferent between the two options is elicited (Bleichrodt et al., 2003; Spencer, 2003). Although in theory both elicitation methods should yield similar results, evidence so far has been mixed (Bleichrodt et al., 2003; Clarke et al., 2003[2]; Spencer, 2003).

[2] This is an unpublished working paper, but was discussed in Bleichrodt et al. (2003). We here only use the information available in that publication.

In this chapter, we report on a new empirical test, set out to address this topic of procedural invariance, in which, unlike previous studies, the TTO scores were also corrected for utility of life duration (discounting) using a new, risk-free method. We introduce the theory that underlies our study in Section 7.2. The experimental setup is described in Section 7.3 and Section 7.4 presents the results. Finally, Section 7.5 discusses the results.

7.2 Background

Preferences over lifetime utility have been represented by an additive utility function over life duration and health quality:

$$V = \sum_{t=j}^{T} \delta_t u(h_t) \tag{1}$$

with $u(h_t)$ denoting a utility function that represents the individual's preferences over health states at each time point t, δ_t the corresponding weight attached to the utility at this point, j the starting period, and T the complete time frame (e.g. Bleichrodt and Gafni, 1996). The conventional TTO procedure makes use of a special case of Equation (1) where $\delta_t = 1$ for each t. This special case is known as the *linear QALY model*, which assumes a linear utility function over life duration. The conventional procedure aims at eliciting the point of indifference between a certain period (n_β) in some health state β followed by death (D), and a shorter period (n_{FH}) in full health (FH), after which again death follows. The elicited indifference can then be represented by the following equation:

$$\sum_{t=0}^{n_\beta} u(\beta) = \sum_{t=0}^{n_{FH}} u(FH) + \sum_{t=n_{FH}}^{n_\beta} u(D) \tag{2}$$

If the utility function over health is normalized so that $u(FH) = 1$ and $u(D) = 0$ we get the following simple expression for $u(\beta)$:

$$u(\beta) = \frac{n_{FH}}{n_\beta} \tag{3}$$

Looking at Equations (2) and (3) it is immediately clear that there are two ways of deriving this value of health state β. One can fix n_β, i.e. the time in health state β, as is common in conventional TTOs, leaving the respondent to specify the period in full health, n_{FH}, making her indifferent between the two options. Alternatively, one might fix the period in full health (n_{FH}) and elicit the number of years in health state β, n_β, again making the respondent indifferent between the two options. According to procedural invariance, both questions, abstracting from biases, should give the same result for $u(\beta)$. Not until recently this feature of TTO gained attention and was used to test the robustness of the TTO method (Bleichrodt et al., 2003; Clarke et al., 2003; Spencer, 2003). The evidence so far is mixed.

Spencer (2003) applied four conventional (10 year period) TTO exercises valuing four different inferior health states. In addition, two alternative TTO exercises were applied in which the period in some inferior health state was fixed and the duration in some worse health state was elicited that yielded indifference between the two scenarios. The results were mixed: only in one question the alternative procedure caused valuations to be significantly lower than those of the conventional procedure, but in the other no significant difference was found. Spencer (2003) moreover provides a useful overview of different biases that influence both procedures and provides qualitative feedback of respondents on the procedures.

Clarke et al. (2003) used a bid procedure to compare the conventional and alternative TTO procedures. Each participant got one bid and had the choice to accept the bid or reject it. They find no evidence of systematic inconsistencies between the procedures.

Finally, Bleichrodt et al. (2003) performed an experiment first performing five conventional TTO exercises all using back pain as the specified health state, but using different time frames (i.e. n_β, ranging from 13 to 38 years). The participants had to return two weeks later, when the alternative TTO procedure was applied using their answers to the conventional TTO (i.e. the elicited n_{FH}) to set the duration of full health. Procedural invariance would then require that the elicited durations with back pain be the same as the originally specified durations, that is, n_β. This way the problem of discounting was circumvented in establishing equivalence of the two procedures. The results indicated significantly lower TTO scores for the alternative procedure compared to the conventional procedure for three out of the five comparisons. For the longest durations no significantly different results were found. Bleichrodt et al. (2003) attribute the differences between the procedures mainly to loss aversion.

Testing for procedural invariance and determining the exact difference in derived health state valuations between the conventional and alternative procedure is not without problems, however, as several biases may lead to differences between the two procedures as highlighted below.

Maximum endurable time

The phenomenon of maximum endurable time means that for severe health states there is a maximum duration people want to live in a health state, after which additional time is valued negatively. Its implication is a higher TTO score for the alternative procedure for higher gauge durations and a lower TTO score for the alternative procedure for lower gauge durations.

Scale compatibility

Scale compatibility may also lead to bias. Scale compatibility refers to the phenomenon that respondents tend to give more weight to the attribute used as the response scale, which is time in the TTO method. This causes an upward bias on TTO scores for both procedures. The magnitude of this bias may differ between the two procedures, depending on the stimuli involved. It should be mentioned that the results of Bleichrodt et al. (2003) are not affected by scale compatibility.

Loss aversion

Loss aversion refers to the fact that people are more sensitive to losses than to gains when viewed from a particular reference point (Tversky and Kahneman, 1991). When we assume that the reference point is the initial health state considered in each question, this will cause a divergence between the two procedures. In the conventional procedure, a loss in longevity occurs when a respondent trades off time, which may cause respondents to be overly reluctant to give up life years, leading to high TTO scores. In the alternative procedure there is a loss in health rather than longevity. In order to compensate for this loss, respondents may require a relatively large gain in life duration to compensate for this loss. As a result, the alternative procedure will generate lower TTO scores (Bleichrodt et al., 2003; Spencer 2003).

Discounting

Finally, since the values of n_{β} and n_{FH} need not be equal between the two TTO procedures, discounting future health is important as it influences the number of years traded or demanded. So, while discounting in itself does not induce a difference between the two procedures, i.e. theoretically both procedures should result in the same TTO score regardless of discounting, it can influence the magnitude of any difference that results from other biases. To what extent the magnitude of this difference is caused by discounting can only be assessed by correcting for it.

We set out to perform a new test of procedural invariance in which we could avoid or correct for two of these biases, i.e. maximum endurable time and utility curvature due to discounting.

7.3 New test

We performed an experimental study comparing the conventional and alternative TTO procedure in the context of a larger study (see Chapters 5 and 6 for more details). A novelty of this study in comparison to the earlier studies is that we correct the TTO scores in both procedures for the influence of discounting[3]. This provides a possibility of looking at the magnitude of the other biases in explaining the difference between the two TTO procedures, without the disturbing influence of discounting. Any difference between the two procedures that remains after correction can be interpreted as a 'net effect' of other biases, in particular loss aversion and scale compatibility.

The health state used in our study was specified as regular back pain.[4] This is a common, easily understandable and non-severely impaired health state.[5] The latter aspect minimizes the influence of the bias of maximum endurable time. We described the health state by the EuroQol 5D terminology. The descriptions were printed on cards and handed to the participants (see Appendix 5A). Further, we

---

[3] Although we use a recently introduced method (Attema et al., 2007), which makes no assumptions regarding the parametric shape of the discount function, correcting TTO scores for discounting is not new (e.g. van der Pol and Roux, 2005). However, this is the first study to do so in the context of examining procedural invariance.

[4] Bleichrodt et al. (2003) successfully used the same health state (albeit using different stimuli) in their test of procedural invariance in a sample of Spanish students. They, however, did not correct TTO scores for discounting in investigating the magnitude of differences between both procedures. While their study could detect procedural invariance regardless of discounting due to its intelligent two-step design, the magnitude of the reported differences due to other biases was influenced by discounting.

[5] The Dutch EQ-5D tariff for this health state (11221 in EQ-5D terms) is 0.811 (Lamers et al., 2005).

stressed to the participants that the health state 'full health' meant they were able to function perfectly on all the five EuroQoL dimensions, irrespective of their age (i.e. the value of the health state was equal to 1). Our experiment entailed two phases, both using regular back pain as the health state of interest. In the first phase of the experiment, we elicited the weighting function for future health (as described in Chapters 5 and 6), used to correct the answers to the TTO questions that were posed in the second phase. In the second phase we used two approaches to value the specified health state. First, we fixed the duration of the health state with back pain (n_β) at 14 (BP14) and 27 years (BP27), respectively, and asked for the number of years in full health (n_{FH}) that they considered equivalent, using an open-ended procedure. Second, in the alternative procedure we fixed the duration in full health (n_{FH}) at 10 (FH10) and 22 (FH22) years, respectively, and asked for the number of years with back pain (n_β) that they considered equivalent by means of an open-ended procedure again.

7.4 Results

Seventy participants, students recruited at the Erasmus University, took part in the experiment. Fourteen participants were eliminated from the study mainly because they did not understand the conventional TTO procedure. The average age of the 56 included respondents was 21.8 years (sd=2.99) and 35.7% was female. Table 7.1 presents the results.

As shown in Table 7.1, the alternative (FH10 and FH22) questions yield substantially lower scores than the conventional (BP14 and BP27) questions, both before and after correction for discounting ($p < 0.01$ for both the raw and the adjusted scores). Correcting for discounting does diminish the difference between the scores of the two procedures considerably. The difference between the raw and adjusted scores is around 0.05 (6%) for the conventional procedure and 0.13 (30%) for the alternative procedure, indicating a substantial difference between the impact of discounting for the two procedures. Longer durations for the alternative procedure explain this difference. It also needs noting that duration is much more influential for the alternative procedure than for the conventional one (i.e. the difference between FH10 and FH22 is much larger than between BP14 and BP27).

The average corrected TTO score for BP14 was still almost 52% higher than for FH10. However, this difference between procedures was much smaller for larger durations, as the average corrected TTO score for BP27 was only about 19% higher than for FH22.

TABLE 7.1. RESULTS.

RAW TTO SCORES				
	BP14	BP27	FH10	FH22
Mean	0.7092	0.7509	0.3711	0.5357
Standard error	0.0247	0.0249	0.0285	0.0240
IQR	0.6429-0.8571	0.7315-0.8889	0.2-0.5	0.3917-0.6332
Range	0.0714-1	0.0741-1	0.1-0.9434	0.22-0.9362
Average value of n	n_{FH}=9.93	n_{FH}=20.28	n_{β}=37.30	n_{β}=46.79
Average number of life years traded/required	4.07	6.73	27.30	24.79
ADJUSTED TTO SCORES				
	BP14	BP27	FH10	FH22
Mean	0.7501	0.8015	0.4935	0.6737
Standard error	0.0249	0.0254	0.0289	0.0251
IQR	0.6911-0.8797	0.6915-0.9353	0.3518-0.6415	0.5178-0.8173
Range	0.0935-1	0.1321-1	0.0862-0.9375	0.2932-0.9839
DIFFERENCE BETWEEN RAW AND ADJUSTED TTO SCORES				
	BP14	BP27	FH10	FH22
Difference	0.0409 (5.8%)	0.0506 (6.7%)	0.1225 (33.0%)	0.1380 (25.8%)

7.5 Discussion

In this study we find significant and substantial differences between the conventional and the alternative procedure to elicit TTO scores, with the conventional procedure invoking considerably higher scores. The correction for discounting we applied indeed diminishes the differences between the procedures, but the differences remain substantial and significant. Compared to earlier studies, our results confirm the findings reported by Bleichrodt et al. (2003). Higher TTO scores for the conventional procedure than for the alternative procedure were also found by Spencer (2003) in one question format, although her estimates were confounded by discounting and maximum endurable time. We moreover confirm the finding of Bleichrodt et al. (2003) that the difference between the two procedures decreases when the specified duration increases. For longer durations than those used here, Bleichrodt et al. (2003) even found no difference between the two procedures.

Of the biases discussed in Section 7.2 only two appear relevant here. We corrected for discounting and avoided maximum endurable time bias by using a relatively mild health problem. As a result, the difference between the two formats relates to scale compatibility and loss aversion for life duration. Both work in the same direction (i.e. higher scores for the conventional TTO), and cannot be disentangled further here. Bleichrodt et al. (2003), however, suggest that loss aversion is the main driver of these differences, leading to an upward bias in the conventional TTO and a downward bias in the alternative TTO. The finding that the difference between the two procedures decreases when the specified duration increases then suggests that people are less loss averse for longer durations. A possible reason for this is that the disutility of the experienced loss (indeed, this loss – moving from healthy to dead – occurs further away in time for conventional TTO procedures) may itself be discounted. Moreover, own expectations regarding duration and quality of life may play a role in such long term TTOs (van Nooten and Brouwer, 2004).[6]

We feel that another factor is relevant in explaining the difference between the two procedures, analogous to the well-known discrepancy between willingness to pay (WTP) and willingness to accept (WTA), the latter normally yielding higher results than the former (e.g. Horowitz and McConnell, 2002). In the WTP approach, respondents need to give up some amount from their own limited budget, which naturally limits their responses. Similarly, in the conventional TTO procedure individuals have a fixed budget of life duration in some non-perfect health state. In "buying" better health their 'willingness to pay' is limited by this budget. The alternative TTO procedure, on the other hand, can be compared to the WTA approach. Then, there is no restriction on the amount of money (in the context of the TTO time) one can ask in compensation for a decline in health. Such a situation without budget restraint can easily lead to higher answers than in case of a budget restraint, resulting in lower TTO scores.

---

[6] Subjective expectations regarding duration of life may have influenced our results as well, since our respondents were relatively young. This may have led them to consider 10 years in full health (the fixed period in FH10) as rather short (in relation to their subjective reference point) and led them to demand relatively many 'unhealthy' life years (the average response in the FH10 question was quite high, i.e. 37.3) in order to regain some of their initial remaining subjective life expectancy. This may also partly explain the large difference between the scores for the FH10 and FH22 questions.

Note that the WTP-WTA gap is also caused by loss aversion (e.g. Knetsch, 1989; Kahneman et al., 1990), but the presence of a possible 'budget constraint' may cause an additional explanation. The open-ended format of our TTO may have added to this discrepancy. Moreover, in case of imprecise preferences, in which some range of answers describes the preferences of the respondent better than one point, one might expect the conventional procedure to lead to an estimate on the low end of this range, while the alternative approaches the range from the other direction and results in a high estimate in that range. Payment scale techniques (e.g. Donaldson et al., 1998) make such relevant ranges explicit, which might be worthwhile investigating in the context of TTO also.

The implication of these findings is that the TTO does not appear consistent across operationalizations. Especially loss aversion appears to play an important role in explaining the difference between the two procedures. Framing TTO as health losses and time gains or vice versa therefore may matter in terms of results. Both framings appear prone to biases and it is difficult to judge which performs better. It is therefore unclear whether we should fix the period of full health or that of an imperfect health state in TTOs. The context of the intervention at stake (preventing health loss or restoring health) may be considered to be important in this respect as well. More fundamentally, we may need to consider whether we can fix the TTO to become less prone to these biases and a more consistent method in deriving health state valuations. For now, it seems unclear that we can.

8 On the (not so) constant proportional tradeoff in TTO[1]

Summary

The preference condition of constant proportional tradeoffs (CPTOs) is necessary for the QALY model to represent preferences over health profiles. The health tariffs used in this model are often estimated by means of the time tradeoff (TTO) method. TTO scores elicited using a particular duration are subsequently attached to health states irrespective of their duration. However, evidence on CPTO so far has been mixed. In this chapter we review this evidence. Further, we use a risk-free method to correct TTO scores for utility curvature and test whether decision makers trade off utility of duration and quality at the same rate irrespective of duration. We find CPTO to be violated for both raw values and utilities. Remarkably, we find higher values for longer durations, contrary to most of the previous studies. We propose a U-shaped relation between TTO scores and duration as a possible explanation for our findings.

---

[1] This chapter is based on Attema and Brouwer (2007c).

8.1 Introduction

The QALY model has become an important model in valuing health benefits. To make the model practical, measurement methods are needed in order to elicit the quality of life weights used in this model. One such method is the TTO method, which is often used to derive (standard) quality of life weights for health states to be used in economic evaluations (e.g. Dolan, 1997; Lamers et al., 2006). The popularity of the TTO, however, is not explained from an absence of methodological problems surrounding it. On the contrary, the TTO has been shown to be prone to several biases and disturbing influences such as loss aversion, scale compatibility and discounting (Bleichrodt, 2002).

One important and necessary assumption for TTO measurements to be consistent with the QALY model is that of constant proportional tradeoffs (CPTOs). CPTO basically requires that the estimated TTO value should be the same for different durations. For example, if in valuing some imperfect health state β using a 10 year TTO people would indicate to be willing to trade off 2 years (that is 20% of total time), then CPTO requires them to give up 2 months when using a 10 *month* TTO or 2 days when using a 10 *day* TTO. The proportion traded should always be equal (i.e. 20%), therefore. Besides a theoretical requirement, CPTO is also practically important when one considers the use of the valuation of health states in economic evaluations and medical decision making: they are attached to such health states regardless of the duration of the health problem, normally. If, therefore, the assumption of CPTO does not hold, health state valuations could be time dependent – that is, health states could be valued differently when their durations differ.

The evidence on the validity of the CPTO assumption is mixed. Some empirical studies found support (e.g. Bleichrodt and Johannesson, 1997; Dolan and Stalmeier, 2003; van der Pol and Roux, 2005), while others rejected it (e.g. Sackett and Torrance, 1978; Stiggelbout et al., 1995), or found mixed results (Bleichrodt et al., 2003). Given the importance of the assumption and the mixed evidence for it, more research in this area seems warranted.

In this chapter we therefore discuss the current evidence regarding CPTO on the basis of a literature review and highlight the role of the utility for life duration in this debate. So far, most studies that found violations of CPTO assumed linear utility (i.e. no discounting or diminishing marginal utility), but it seems implausible that their participants would satisfy that assumption. Therefore, if one would have corrected for utility curvature, these respondents might have satisfied CPTO in terms of utilities after all. That is, TTO scores corrected for utility curvature may still be the same for different durations, despite the fact that 'raw' TTO scores vary with duration. It is important to investigate this possibility, because it might indicate that the QALY model does hold in a more general form, and that only the utility for life duration has to become less restrictive. We present the results of an experiment to test the CPTO assumption, in which we used both raw TTO scores and TTO scores that were corrected for utility curvature. For the utility correction of the TTO values, we use the risk-free utility for life duration elicitation method of Attema et al. (2007). Its advantages are that it does not need to make specific parametric assumptions about the utility function and that it is not influenced by biases due to probability weighting and the inclusion of the problematic outcome death. (See Chapter 5 for a discussion.)

This chapter is organized as follows. We describe the theoretical background of CPTO in Section 8.2. Then, we review the existing literature that tested the CPTO assumption in Section 8.3. Our experimental test, followed by a presentation of the results, is described in Section 8.4. Finally, Section 8.5 discusses the results and provides a possible explanation for our findings in the form of a generalized relationship between duration and tradeoffs.

8.2 TTO and constant proportional tradeoffs

As indicated, the TTO method is based on the QALY model. The QALY model is a common way to describe preferences over lifetime utility and represents these preferences by the following additive utility function over life duration and health quality (e.g. Miyamoto and Eraker, 1988):

$$V = \sum_{t=j}^{T} \delta_t u(h_t) \qquad (1)$$

with $u(h_t)$ a utility function that represents the individual's preferences over health states at each time point t, δ_t denoting the corresponding weight attached to the utility at this point, j the starting period, and T the complete time frame.

Often, TTOs assume linear utility, so that $\delta_t = 1$ for each t in Equation (1). The model then simplifies to the *linear QALY model*, where equal weight is assumed to be attached to all utilities regardless of their timing. Then, the utility of a health state β can be elicited by asking the respondent to give some period in full health, followed by death, which makes him indifferent to a stated period in health state β, also followed by death. As a result, the elicited indifference can be represented by the following equation:

$$n_\beta * u(\beta) = n_{FH} * u(FH) + (n_\beta - n_{FH}) * u(D) \tag{2}$$

with *FH* representing full health and *D* death, n_β is the stated period in health state β, and n_{FH} is the elicited period in full health. If the utility function over health is normalized so that $u(FH) = 1$ and $u(D) = 0$ we get the following simple expression for $u(\beta)$:

$$u(\beta) = \frac{n_{FH}}{n_\beta} \tag{3}$$

Clearly, this expression only holds if the utility over life duration is indeed linear and there is no reallocation of lifetime consumption due to the smaller duration in perfect health.[2] Pliskin et al. (1980) gave an axiomatic derivation of a particular version of the QALY model for constant health profiles, which includes the linear QALY model. They proved that preferences can be represented by this model if utility independence and CPTO hold. *CPTO* means that for each health state β there exists a number $q \geq 0$, such that n_β years in β is equivalent to qn_β (= n_{FH} in Equation (3)) years in full health (*FH*), i.e. $(\beta, n_\beta) \sim (FH, qn_\beta)$. In other words,

---

[2] The latter concern will probably not cause a large bias (Dolan and Jones-Lee, 1997).

individuals are willing to give up the same proportion (1-q) of lifetime irrespective of its duration (n_β).

Pliskin et al. (1980) also showed that if utility independence and CPTO hold, the utility function for life duration has to be linear or a power function. However, individuals may have a utility of life duration function that does not belong to the power family, but instead to some other parametric family. In that case, CPTO does not need to be confirmed, but it may very well be that such an individual does obey the assumption of CPTO in terms of *utilities of life years* instead of ordinary life years. That is, for each health state β there may exist a number $q \geq 0$, such that the utility of n_β years in β is equivalent to the utility of n_β years in full health multiplied by q. In other words, individuals may be willing to give up the same proportion (1-q) of utility of lifetime irrespective of its duration (n_β). Consequently, the QALY model of Pliskin et al. (1980) may still hold, albeit with less restrictions on the shape of the utility function for life duration. When the utility function for life duration is exponential instead of a power, for example, CPTO may hold in terms of utilities but not in terms of ordinary life years. Therefore, testing this form of CPTO requires the correction of TTO scores for utility curvature. Recently, Attema et al. (2007) have pointed out a new way of correcting for this curvature, using a risk-free elicitation method. Attema and Brouwer (2007a) used this method to adjust TTO scores.

CPTO can be violated due to other reasons as well (see e.g. Bleichrodt, 2002), like loss aversion (i.e. the phenomenon that people are more sensitive to losses than to gains when viewed from a particular reference point—Tversky and Kahneman (1991)), and maximum endurable time (i.e. the fact that some bad health states can only be endured during some period of time after which its value becomes negative). Depending on the magnitude and direction of these biases, CPTO may be violated in both directions or the biases may cancel out so that CPTO is not violated on the aggregate – or may mistakenly be perceived as not being violated if the violation itself is balanced by other biases. A better understanding of the validity of the assumption of CPTO therefore depends on a better understanding of the magnitudes of these biases and correcting for them as far as possible.

In the conventional TTO procedure, loss aversion may cause respondents to be overly reluctant to give up life years, leading to relatively high utility scores. Maximum endurable time will lead to higher values for bad health states for small durations, because for longer durations extra time in that health state will be valued negatively. While loss aversion is likely to be present in all TTO valuations, the presence of MET depends on the health state valued. Moreover, the influence of utility curvature, for instance caused by discounting, is also present in normal TTO valuations, but its influence will likely vary with the time horizon chosen and can be corrected for. In the next section, we will highlight the existing evidence regarding CPTO. In doing so, we will indicate whether the performed studies corrected for utility curvature or used health states that could lead to MET bias.

8.3 Empirical evidence on CPTO

As noted before, the evidence about the empirical validity of the CPTO assumption is mixed. This section highlights the existing evidence on the validity of the CPTO assumption. Table 8.1 presents the empirical studies regarding CPTO, summarizes their main results (in terms of significance of the difference between small and long durations at the 5% level) and indicates whether or not they corrected for utility of life duration curvature and whether or not they used a health state that may be susceptible to maximum endurable time. The table emphasizes the amount of variation in results. Some studies confirm CPTO, but most studies reject it. These violations, however, are not easily interpretable since CPTO is violated in both directions, i.e. sometimes the proportion traded is relatively small for shorter durations and sometimes the proportion traded is larger for shorter durations, compared to longer durations. The finding that the TTO values of some health state are higher for short durations (i.e. less life years are being traded in that case) is somewhat more common.

TABLE 8.1. OVERVIEW OF CPTO STUDIES.

Study	**Life years used**	**MET health state**	**Utility correction**	**Results**	**Sample**
Sackett and Torrance (1978)	3 months 8 years Life expectancy	Mixed	No	TTO L < TTO S	General population Patients
Pliskin et al. (1980)	5 years 15 years	No	Yes	TTO L = TTO S	Pilot sample
Miyamoto and Eraker (1988)	1, 2, 15, 16, 20, 24	No	No	TTO L < TTO S (p-values not reported)	Patients
Hall et al. (1992)	10% of LE 50% of LE Life expectancy	Mixed	No	TTO L = TTO S	Women 40-70 50% patients
Cook et al. (1994)	1 year 12 years	No	No	TTO L = TTO S	Patients
Stiggelbout et al. (1995)	3, 10, 15 years 3, 5, 10 years 5, 20, LE	No	No	TTO L < TTO S	Patients
Stalmeier et al. (1996)	5, 10, 25, 50	Yes	Yes	TTO L = TTO S Except for t=5	Students
Stalmeier et al. (1997)	5, 10, 25, 50	Yes	No	TTO L = TTO S TTO L > TTO S	Students
Bleichrodt and Johannesson (1997)	10 and 30 years	No	No	TTO L = TTO S	Students
Unic et al. (1998)	5, 10 and higher	No	No	TTO L > TTO S	Healthy women

Study	Life years used	MET health state	Utility correction	Results	Sample
Kirsch and McGuire (2000)	2 and 10 years	Mixed	No	TTO L = TTO S TTO L < TTO S	General population
Martin et al. (2000)	5, 10, 15 years	No	Yes	TTO L < TTO S	Older cardiovascular disease patients (mean age 61)
Stalmeier et al. (2001)	10 and 20 years	Yes	No	TTO L < TTO S (p-values not reported)	Students Patients
Bleichrodt et al. (2003)	13, 19, 24, 31, 38 years	No	No	TTO L = TTO S TTO L > TTO S	Students
Dolan and Stalmeier (2003)	10 and 20 years	Yes	No	TTO L < TTO S	Students
Van der Pol and Roux (2005)	20 and 50 years	No	Yes	TTO L = TTO S	Students

Sackett and Torrance (1978) used a small, an intermediate and a long duration for a general population sample and a patient sample, and report higher TTO values for short durations. Miyamoto and Eraker (1988) reported evidence of no trade at all for durations smaller than 1 year for 25% of the respondents, whereas these people did trade off time for longer durations. Therefore, TTO values were higher for small durations than for long durations. Hall et al. (1992) compared three different durations, depending on life expectancy of the respondents. No violations of CPTO were found. Cook et al. (1994) interviewed patients. Their TTO values were not significantly different between a duration of 1 year and one of 12 years. Stiggelbout et al. (1995) used small and intermediate durations and found a violation of CPTO with TTO values for small durations being higher than those for long durations. Bleichrodt and Johannesson (1997) used an intermediate and a long duration. They found no violation of CPTO at the aggregate level.

Stalmeier et al. (1997) used bad health states to test for maximum endurable time. They cannot reject CPTO when comparing durations of 10 years and longer, but do find significantly lower TTO values for the 5 year horizon. Unic et al. (1998) estimated TTO values for several durations in a sample of healthy women and found lower values for shorter durations. Kirsch and McGuire (2000) compared a small and an intermediate duration and found mixed evidence. They reported higher TTO values for the short duration for bad health states, but no significant differences for moderate health states. They attribute this to respondents who value additional time in a bad health state as worse than death after some duration (i.e. maximum endurable time). Stalmeier et al. (2001) and Dolan and Stalmeier (2003) found smaller TTO values for higher durations when comparing two intermediate durations in a severe health state. This may again have been caused by maximum endurable time. Bleichrodt et al. (2003) used five different durations that were no multiples of 5 so that they were not very susceptible to a proportional heuristic. These durations were of an intermediate and long-term nature. They found higher TTO values for high durations than short durations for one procedure, indicating that people are willing to trade off relatively less life years for higher amounts. For another procedure, they could, however, not reject CPTO.

We found four studies that corrected for utility of life duration curvature. Pliskin et al. (1980) used a certainty equivalence (CE) question to correct for utility of life duration and found no violation at the aggregate level. However, their sample was very small. Stalmeier et al. (1996) found no violation of CPTO. They also corrected for utility curvature by means of the CE method and estimated several parametric models. Martin et al. (2000) used three short and short-intermediate durations and corrected for utility curvature by means of CE questions. In a sample of cardiovascular patients, they found smaller TTO values for higher durations. Van der Pol and Roux (2005) compared TTO scores for a long-intermediate duration (20 years) and a very long duration (50 years). Further, they corrected for discounting by means of one discounting question. They found no violation of CPTO, neither for unadjusted nor for individually adjusted scores.

To summarize, sixteen empirical studies of CPTO were found. Six of these did not reject CPTO, six found lower TTO values for higher durations, one found the opposite result, and three found mixed results. There is no clear influence of correcting for discounting nor is there a clear influence of MET.[3] It appears difficult therefore to derive any definite answers from the literature regarding CPTO. Most evidence points towards higher values for short durations, yet all but one of these studies do not correct for discounting, which can strongly influence results, given the time horizons chosen. It appears that more evidence is required

---

[3] It needs noting that TTO studies generally vary quite strongly in terms of designs (Arnesen and Trommald, 2005). Such variations obviously can also hamper direct comparisons of results of the studies listed in Table 8.1.

in order to better understand the relationship between health state duration and valuation.

8.4 Testing CPTO while correcting for utility curvature and avoiding MET

We performed an experimental study testing the CPTO assumption in the context of a larger study (see Chapters 5 and 6 for more details). Our experiment entailed two phases. In both phases we used regular back pain as the health state of interest. Back pain is a common, easily understandable and non-severe health state. The latter aspect minimizes the influence of the bias of maximum endurable time. We described the health state by the EuroQol 5D terminology. The descriptions were printed on cards and handed to the participants (see Appendix 5A). Further, we stressed to the participants that the health state 'full health' meant they were able to function perfectly on all the five EuroQoL dimensions, irrespective of their age. In the first phase of the experiment, we elicited the weighting function for future health, used to correct the answers to the TTO questions that were posed in the second phase.

The first phase was based on the notion that if utility of life duration is nonlinear, the estimate obtained by Equation (3) will be biased. Rather, one would then need to use Equation (1). Without restrictions on δ_t (and using the notation introduced above) the indifference implied by the TTO under the QALY model would give the following equation:

$$\sum_{t=0}^{n_\beta} \delta_t u(\beta) = \sum_{t=0}^{n_{FH}} \delta_t \qquad (4)$$

Eliciting $u(\beta)$ thus requires the elicitation of the weights δ_t. Attema et al. (2007) recently proposed a new risk-free elicitation method to do so. In short, the method presents a participant with two health scenarios. In the first scenario the participant first is in a good health state (*g*). After some time, *t*, she moves to a worse health state (*h*) for the remainder of the total time period P. In the second scenario, the participant first is in the worse health state (*h*) and at time *t* moves to the better health state (*g*) for the remainder of P. The value of *t* is elicited that makes the participant indifferent between these scenarios. This value indicates the point where the period before *t* yields as much utility as the period after *t*. When *t* is smaller than the midpoint of the period P, this indicates concavity of the utility function over life duration. Then, as a result, raw TTO values will be biased downwards and correcting for this concavity results in higher utility scores. More detailed information about the exact shape of the utility function can be obtained by repeating this procedure (using the first estimate of *t* as input in the next exercise, etc). (See Chapter 5 for details).

In the second phase of the experiment we used two approaches to value the specified health state. First, in a conventional procedure, we fixed the duration of the health state with back pain (n_β) at 14 (BP14) and 27 years (BP27), respectively, and asked for the number of years in full health (n_{FH}) that they considered equivalent. Second, in an alternative procedure, we fixed the duration in full health (n_{FH}) at 10 (FH10) and 22 (FH22) years, respectively, and asked for the number of years with back pain (n_β) that they considered equivalent.

Results

Seventy participants were recruited and were paid a fixed amount of €12.50 to join the experiment. The participants were students from different faculties of the Erasmus University Rotterdam. Fifty-six participants were included in the analyses. The other 14 participants were eliminated from the sample because they had at least one inconsistent answer or had not understood the utility elicitation part. The average age of the 56 included respondents was 21.8 years (sd=2.99) and 35.7% was female.

In Table 8.2 we present some summary statistics concerning the raw and corrected TTO scores.[4] The difference between the raw and adjusted values is around 0.05 (6%) for the BP questions and 0.13 (30%) for the FH questions (see last row of Table 8.2).

[4] These results were used to test the procedural invariance of TTO in Attema and Brouwer (2007b).

TABLE 8.2. TTO SCORES.

Raw scores	**BP14**	**BP27**	**FH10**	**FH22**
Mean	0.7092	0.7509	0.3711	0.5357
Standard error	0.0247	0.0249	0.0285	0.0240
Interquartile range	0.6429-0.8571	0.7315-0.8889	0.2-0.5	0.3917-0.6332
Range	0.0714-1	0.0741-1	0.1-0.9434	0.22-0.9362
Average number of years required/sacrificed	4.07	6.73	27.30	24.79
Adjusted scores				
Mean	0.7501	0.8015	0.4935	0.6737
Standard error	0.0249	0.0254	0.0289	0.0251
Interquartile range	0.6911-0.8797	0.6915-0.9353	0.3518-0.6415	0.5178-0.8173
Range	0.0935-1	0.1321-1	0.0862-0.9375	0.2932-0.9839
Difference between BP14 and BP27: 0.0417 (5.9%, $p < 0.01$) Difference between adjusted BP14 and adjusted BP27: 0.0514 (6.9%, $p < 0.01$) Difference between FH10 and FH22: 0.1646 (44.4%. $p < 0.01$) Difference between adjusted FH10 and adjusted FH22: 0.1802 (36.5%, $p < 0.01$)				

In order to test the CPTO assumption, we compared the small and the long duration for both elicitation procedures. For both raw and adjusted TTO scores, CPTO was rejected, with the score being higher for longer than for smaller durations ($p < 0.01$). This finding is in contrast with most of the aforementioned studies.

In the FH questions, our results seem to be caused partly by the large fraction of respondents (23.2%) that gave the same answer to FH10 and FH22. In the BP questions, many respondents (33.9%) answered as if using a proportional heuristic, i.e. their answer to the second question was twice the amount of their answer to the first question. Because the input of the second question was somewhat lower than twice the amount of the first question (27<2*14), this resulted in a higher raw TTO score for a longer duration for these respondents. We therefore repeated the analysis excluding these respondents, which still yielded a significant difference between the different durations. For BP14 and BP27, CPTO is still rejected in the same direction ($p < 0.02$). For the FH questions, FH22 also still yields higher TTO scores than FH10, both when excluding proportional heuristic respondents and when excluding respondents who gave the same answer to both questions ($p < 0.01$. For the BP questions there were no respondents giving the same answers to both questions). As a result, for the alternative procedure, there is again a violation of CPTO in the opposite direction of most of the earlier found violations.

Summarizing, our results indicate that correcting for utility curvature and avoiding MET does not seem to be sufficient to restore the validity of the assumption of CPTO.

8.5 Discussion

What can we infer from this overview and study other than that we succeed in adding to the confusion regarding constant proportional tradeoff? We believe some important observations need to be made.

First, the review of the literature shows that violations of CPTO are common. Though often the violation causes shorter duration to result in a decreased willingness to trade and therefore higher health state valuations, the opposite has also been shown. The reviewed studies differ in many respects, including the time horizon chosen and whether a correction for utility curvature has been applied. Not many studies do the latter. Of the four that did, three find no violation of CPTO, while one finds that shorter durations result in higher valuations (Martin et al., 2000). The fact that the latter study used relatively old patients (average age of 61) in their study may have influenced results, not only because of the way they view health problems, but also because of the fact that their subjective life-expectancy may have been less than the projected ones. (See van Nooten and Brouwer (2004) on how this could bias results.) Such differences between the studies make it difficult to derive general conclusions from the existing evidence.

The present study was clearly small and the sample consisted of students, hampering generalization. Still, we found a robust violation of CPTO for both raw and corrected TTO scores in our sample. Remarkably, this violation is in the opposite direction of most of the previously found violations of CPTO and the only study correcting for utility curvature to find such a violation of CPTO. We also found that the magnitude of the violation was much smaller for the conventional TTO procedure (fixing time in an imperfect health state) than for the alternative one (fixing the period in full health). The latter was also found by Bleichrodt et al. (2003) and stresses the importance of other biases and influences in deriving the violation of CPTO.

Our results concerning the utility of life duration (see Chapters 5 and 6) are in agreement with the violation of CPTO for raw scores. We find evidence against a constant relative risk attitude for remaining lifetime, which is the property characterizing power utility functions. Instead the data seem to be more in line with a constant absolute risk attitude, corresponding to exponential utility functions. However, after correcting for utility curvature, CPTO is still rejected, which indicates a more fundamental rejection of the QALY model. It seems that individuals do not trade off utility of life duration for health status at a constant rate, but instead at a rate that depends on the duration involved. For relatively long durations, like the ones used in our study, the amount of years traded is relatively low also after correction for utility curvature. Given this finding, the plausibility of relatively high TTO values for very short durations (who would trade off two days to avoid low back pain when having only ten days left to live?) and the diverse violations of CPTO reported in the literature (which indeed must be related to the fact that TTO results vary strongly between studies as reported by Arnesen and Trommald (2005)), it is interesting to hypothesize on the shape of this relationship between duration and tradeoffs.

Given the importance of loss aversion in the TTO (e.g. Bleichrodt, 2002; Bleichrodt et al., 2003), we hypothesize that a possible explanation for the variation in findings and therefore for a general relationship between health state duration and health state valuation in TTO is driven by this bias. In a conventional TTO with a 'small' duration, loss aversion and scale compatibility may relate especially to the amount of time left to live and may be stronger for smaller time horizons (durations). Loss aversion then causes respondents to be overly reluctant to give up life years, leading to relatively high utility scores. For 'long' durations, on the other hand, the absolute amount of years sacrificed may become dominant in the tradeoff, i.e. the reference point of the respondents changes, with people being reluctant to trade off more than some absolute amount of time. Thus, the absolute amount of time remaining is leading when the TTO uses short durations and the absolute amount of time sacrificed is leading for longer durations. The result will be that individuals give up fewer years for short and long durations, and less driven by these considerations in between these two points, causing TTO values to be a U-shaped function of duration. Future research testing multiple durations using a within-subject design could formally test this hypothesis.

It is important to stress that the use of the word ‘bias’ can be misleading. Not all deviations from CPTO need to be ‘biases’ in the sense that the TTO method causes some systematic misrepresentation of real preferences. The biases discussed here may simply reflect genuine and even plausible preferences, which are simply not adequately captured in the current QALY model. For instance, desiring some minimal remaining length of life seems a common and plausible preference, which can affect health state valuations elicited through TTO. Of course, such time dependency in valuations has important implications for deriving health state valuations and for the practice of economic evaluations, if we would want to reflect such preferences.

For now, it seems that the QALY model may be too simple, that there is indeed no constant proportional tradeoff of life years for health quality and, therefore, that health state valuations may depend on the duration of these health states.

9 Conclusion

This chapter discusses the main conclusions drawn in this thesis. The conclusions will be made along the lines of the research questions defined in the introduction.

9.1 Time preference

Chapter 2 reviewed the existing literature on time preference and showed that the violations of the constant discounting model are extensive. It also discussed their implications for medical decision making and presented several examples of applications to health-related behavior. It made clear that incorporating the observed violations of the standard model into new models is able to explain anomalous health-related behavior and can be exploited to improve policy recommendations.

Chapters 3 and 4 contributed to the empirical literature on time preference by introducing new measurement methods and performing experimental measurements of discounted utility with these methods. In Chapter 3 a new method to measure intertemporal preferences was proposed, where first utility of money is elicited in a nonparametric way. Moreover, the method elicits utility in an intertemporal domain, so that a uniform setting is used throughout the entire measurement process. Thereafter, time preference can be elicited, correcting for utility curvature as determined in the first stage. The method was subsequently tested in an experiment. It turned out that intertemporal utility was concave for gains and convex for losses, consistent with a hypothesis of Loewenstein and Prelec (1992). However, utility curvature had not much influence on time preferences. It did lower the gain-loss asymmetry somewhat, but the difference in discount factors between gains and losses remained significant. Another interesting result is that I found this asymmetry even though I used a neutral frame. Therefore, I rejected Shelley's (1993) conjecture that the gain-loss asymmetry can be explained by a framing effect. Further, I found the generalized hyperbolic discounting model of Loewenstein and Prelec (1992) to describe the data significantly better than the constant discounting model, whereas other hyperbolic discounting models gave a similar fit as the constant discounting model. An implication thereof is that impatience is decreasing monotonically over time, and, hence, hyperbolic discounting is not merely caused by an immediacy effect as in quasi-hyperbolic discounting.

A way to measure the degree of time inconsistency, i.e. the deviation from stationarity, without needing information about the utility function for money, was proposed in Chapter 4. This measure was subsequently used in an experiment. Violations of both constant and hyperbolic discounting were found and, instead, discounting was increasing over time, contrary to most of the previous evidence. These results make clear that observed time preferences depend heavily on the elicitation procedure. The experiment of Chapter 3 used a choice procedure and expressed delay in terms of months and years, whereas Chapter 4 made use of a matching procedure and expressed delay in terms of months only. Another important difference concerns the response scale. Chapter 3 had money as response scale, whereas Chapter 4 had time as response scale. More research on the influence of the procedure is therefore warranted.

Chapter 5 investigated time preference for future lifetime. It proposed a risk-free method for measuring the utility of life duration. The advantages of this method over existing methods are that it is not distorted by probability weighting and that is does not need the inclusion of the problematic outcome death. The results of a questionnaire confirmed that respondents find this method easier to answer than both the certainty equivalence method and the tradeoff method, which both measure utility under risk. Utility of life duration was measured in an experiment and compared regarding the three methods. The certainty equivalence method yielded more concave utility than the risk-free method, but this difference was no longer significant after correction for probability weighting. The results of the tradeoff method, which is not distorted by probability weighting, did not differ significantly from those of the risk-free method. It therefore seems that the risk-free method is able to provide a reliable measure of utility and is easy to apply for practical purposes.

Another remarkable finding was that utility could be described better by an exponential function than by the popular power function. These results lend support to a constant *absolute* risk posture over life years instead of a constant *relative* risk posture. This is in contrast to other studies that did not find this result (e.g. Abellan-Perpinan et al., 2006), and is some evidence against the QALY model proposed in the seminal study of Pliskin et al. (1980), because that model requires a linear or power utility for life duration function. An exponential utility for life duration function, on the other hand, has the interesting property that it corresponds with constant discounting of future life years, which is common practice in health economics.

9.2 Universality of utility

In addition to measuring time preference, the methods developed in this thesis gave the possibility to compare utility in different domains. Chapter 3 compared utility of money elicited in an intertemporal domain to previous findings on utility elicited in risk and uncertainty domains. The findings were rather similar, indicating a universal concept of utility.

Chapter 5 tested whether utility of life duration in a certainty domain differed from utility of life duration in a risky domain. No significant differences were found when correcting for probability weighting. Keeping in mind that probability weighting is a bias that is distinct from utility curvature, this finding is again evidence in favor of universality of utility. These results have important implications. For instance, they support the transferability of utility through different domains and as such support the common practice among health scientists to apply TTO scores (time domain) and standard gamble utilities (risk domain) in economic evaluations (welfare economic domain). Moreover, these results reject the common view in economics that utility is context dependent.

9.3 Time tradeoff method

This thesis has applied the risk-free method of Chapter 5 to TTO valuations in order to investigate the role of the utility of life duration in the TTO method. First, in Chapter 6, I explained how to correct TTO scores for utility of life duration curvature with the risk-free method and estimated the size of this correction. Due to the concave shape of the utility functions, the corrected TTO scores were significantly higher than the uncorrected ones. The magnitude of this difference was approximately 0.05 (6%).

Chapter 7 dealt with procedural invariance of the TTO method. There I considered the influence of utility for life duration on the disparity between two TTO procedures. It was found that correcting for utility of life duration diminishes this disparity, although a large and significant gap remains. This is probably caused by loss aversion.

Finally, Chapter 8 considered the constant proportional tradeoffs property. The existing evidence on this property was reviewed and a new test was performed that investigated whether utilities of life years were traded off in a constantly proportional way. CPTO turned out to hold neither for ordinary life years nor for utilities. This result implies that the QALY model for decision making in health needs reconsideration as a descriptive model for individual preferences over health outcomes.

References

Abdellaoui, M. 2000. Parameter-free elicitation of utility and probability weighting functions. *Management Science* **46 (11)** 1497-1512.

Abdellaoui, M., A.E. Attema, H. Bleichrodt. 2006. Intertemporal tradeoffs for gains and losses: An experimental measurement of discounted utility. *Submitted.*

Abdellaoui, M., C. Barrios, P.P. Wakker. 2007. Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory. *Journal of Econometrics* **138** 356-378.

Abdellaoui, M., F. Vossmann, M. Weber. 2005. Choice-based elicitation and decomposition of decision weights for gains and losses under uncertainty. *Management Science* **51** 1384-1399.

Abellan-Perpinan, J.M., J.L. Pinto, I. Mendez, X. Badia-Llach. 2006. Towards a better QALY model. *Health Economics* **15 (7)** 665-676.

Ahlbrecht, M., M. Weber. 1997. An empirical study on intertemporal decision making under risk. *Management Science* **43** 813-826.

Ainslie, G. 1975. Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin* **82** 463-496.

Airoldi, M., D. Read, S. Frederick. 2005. Present bias and dynamic inconsistency: a longitudinal study. *Work in progress.*

Andersen, S., G.W. Harrison, M.I. Lau, E.E. Rutström. 2006. Eliciting risk and time preferences. *Working paper.*

Antonides, G., S.R. Wunderink. 2001. Subjective time preference and willingness to pay for an energy-saving durable good. *Zeitschrift für Sozialpsychologie* **32 (3)** 133-141.

Arnesen, T., M. Trommald. 2005. Are QALYs based on the time trade-off comparable? A review of TTO-methodologies. *Health Economics* **14 (1)** 39-53.

Arrow, K.J. 1951. Alternative approaches to the theory of choice in risk-taking situations. *Econometrica* **19** 404-437.

Attema, A.E. 2006. Measuring proneness to arbitrage using only pencil and paper. *Medium Econometrische Toepassingen* **14 (4)** 2-6.

Attema, A.E., H. Bleichrodt, K.I.M. Rohde, P.P. Wakker. 2006. Time-tradeoff sequences for quantifying and visualizing the degree of time inconsistency, using only pencil and paper. *Working paper, Erasmus University Rotterdam.*

Attema, A.E., H. Bleichrodt, P.P. Wakker. 2007. A new method for measuring the utility of life duration. *Working paper, Erasmus University Rotterdam.*

Attema, A.E., W.B.F. Brouwer. 2007a. The correction of TTO-values for utility curvature using a risk-free utility elicitation method. *Submitted.*

Attema, A.E., W.B.F. Brouwer. 2007b. Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO-measurement. *Health Economics, forthcoming.*

Attema, A.E., W.B.F. Brouwer. 2007c. On the (not so) constant proportional tradeoff in TTO-measurement. *Submitted.*

Baker, F., M.W. Johnson, W.K. Bickel. 2003. Delay discounting in current and never-before cigarette smokers: similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology* **112** 382-392.

Barsky, R.B., F.T. Juster, M.S. Kimball, M.D. Shapiro. 1997. Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics* **112** 537-579.

Becker, G.S., K.M. Murphy. 1988. A theory of rational addiction. *Journal of Political Economy* **96 (4)** 675-700.

Bénabou, R., J. Tirole. 2002. Self-confidence and personal motivation. *Quarterly Journal of Economics* **117** 871-915.

Benhabib, J., A. Bisin, A. Schotter. 2006. Present bias, quasi-hyperbolic discounting, and fixed costs. *Working paper.*

Benzion, U., A. Rapoport, J. Yagil. 1989. Discount rates inferred from decisions: An experimental study. *Management Science* **35** 270-284.

Bernheim, B.D., J. Skinner, S. Weinberg. 2001. What accounts for the variation in retirement wealth among U.S. Households? *American Economic Review* **91** 832-857.

Bickel, W.K., M.W. Johnson. 2003. Delay discounting: A fundamental behavioral process of drug dependence. G.F. Loewenstein, D. Read, R.F. Baumeister, eds. *Time and decision.* Russell Sage Foundation, New York, 419-440.

Bickel, W.K., A.L. Odum, G.J. Madden. 1999. Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers. *Psychopharmacology* **146** 447-454.

Bleichrodt, H. 2001. Probability weighting in choice under risk: An empirical test. *Journal of Risk and Uncertainty* **23** 185-198.

Bleichrodt, H. 2002. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics* **11** 447-456.

Bleichrodt, H., J.M. Abellan-Perpinan, J.L. Pinto, I. Mendez. 2007. Resolving inconsistencies in utility measurement under risk: Tests of generalizations of expected utility. *Management Science* **53** 469-482.

Bleichrodt, H., A. Gafni. 1996. Time preference, the discounted utility model and health. *Journal of Health Economics* **15** 49-66.

Bleichrodt, H., M. Johannesson. 1997. The validity of QALYs: An empirical test of constant proportional tradeoff and utility independence. *Medical Decision Making* **17** 21-32.

Bleichrodt, H., M. Johannesson. 2001. Time preference for health: A test of stationarity versus decreasing timing aversion. *Journal of Mathematical Psychology* **45** 265-282.

Bleichrodt, H., J.L. Pinto. 2000. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science* **46** 1485-1496.

Bleichrodt, H., J.L. Pinto. 2005. The validity of QALYs under non-expected utility. *Economic Journal* **115** 533-550.

Bleichrodt, H., J.L. Pinto, J.M. Abellan-Perpinan. 2003. A consistency test of the time trade-off. *Journal of Health Economics* **22** 1037-1052.

Bleichrodt, H., J.L. Pinto, P.P. Wakker. 2001. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science* **47** 1498-1514.

Bommier, A. 2005. Life-cycle theory for human beings. *Working paper.*

Bommier, A. 2006. Uncertain lifetime and intertemporal choice: risk aversion as a rationale for time discounting. *International Economic Review* **47 (4)** 1223-1246.

Bostic, R., R.J. Herrnstein, R.D. Luce. 1990. The effect on the preference reversal of using choice indifferences. *Journal of Economic Behavior and Organization* **13** 193-212.

Bretteville-Jensen, A.L. 1999. Addiction and discounting. *Journal of Health Economics* **18** 393-407.

Brocas, I., J.D. Carrillo. 2000. The value of information when preferences are dynamically inconsistent. *European Economic Review* **44** 1104-1115.

Brocas, I., J.D. Carrillo. 2001. Rush and discounting under hyperbolic discounting and interdependent activities. *Journal of Risk and Uncertainty* **22** 141-164.

Cairns, J.A. 1992. Health, wealth and time preference. *Project Appraisal* **7 (1)** 31-40.

Cairns, J.A. 1994. Valuing future benefits. *Econometrics and Health Econometrics* **3** 221-229.

Cairns, J.A., M.M. van der Pol. 1997a. Saving future lives: a comparison of three discounting models. *Health Economics* **6 (4)** 341-350.

Cairns, J.A., M.M. van der Pol. 1997b. Constant and decreasing timing aversion for saving lives. *Social Science and Medicine* **45 (11)** 1653-1659.

Cairns, J.A., M.M. van der Pol. 2000. The estimation of marginal time preference in a UK-wide sample (TEMPUS) project. *Health Technology Assessment* **4** 1-73.

Camerer, C.F. 1995. Individual decision making. J. Kagel, A. Roth, eds. *The handbook of experimental economics*. Princeton University Press, Princeton, NJ, 587-703.

Camerer, C.F., R.M. Hogarth. 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* **19** 7-42.

Chapman, G.B. 1996a. Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning Memory and Cognition* **22** 771-791.

Chapman, G.B. 1996b. Expectations and preferences for sequences of health and money. *Organizational Behavior and Human Decision Processes* **67** 59-75.

Chapman, G.B. 1998. Sooner or later: the psychology of intertemporal choice. D.L. Medin, ed. *The psychology of learning and motivation* **38** 83-113.

Chapman, G.B. 2000. Preferences for improving and declining sequences of health outcomes. *Journal of Behavioral Decision Making* **13** 203-218.

Chapman, G.B. 2003. Time discounting of health outcomes. G.F. Loewenstein, D. Read, R.F. Baumeister, eds. *Time and decision.* Russell Sage Foundation, New York, 395-417.

Chapman, G.B. 2005. Short-term cost for long-term benefit: time preference and cancer control. *Health Psychology* **24 (4)** S41-S48.

Chapman, G.B., N.T. Brewer, E.J. Coups, S. Brownlee, H. Leventhal, E.A. Leventhal. 2001. Value for the future and preventive health behavior. *Journal of Experimental Psychology: Applied* **7 (3)** 235-250.

Chapman, G.B., E.J. Coups. 1999. Time preferences and preventive health behavior: acceptance of the influenza vaccine. *Medical Decision Making* **19** 307-314.

Chapman, G.B., A.S. Elstein. 1995. Valuing the future: temporal discounting of health and money. *Medical Decision Making* **15** 373-386.

Chapman, G.B., R. Nelson, D.B. Hier. 1999. Familiarity and time preferences: decision making about treatments for migraine headaches and Crohn's disease. *Journal of Experimental Psychology: Applied* **5 (1)** 17-34.

Clarke, P., J. Wolstenhome, K. Johnston. 2003. Time trade-off or time gain?: It all depends on your point of reference. *Working paper, University of Oxford.*

Coller, M., G.W. Harrison, E.E. Rutström. 2005. Does everyone have quasi-hyperbolic preferences? *Working paper.*

Cook, J., J. Richardson, A. Street. 1994. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. *Health Economics* **3 (3)** 157-168.

Cropper, M.L., S.K. Ayded, P.R. Portney. 1992. Rates of time preference for saving lives. *American Economic Review* **82** 469-472.

Cropper, M.L., S.K. Ayded, P.R. Portney. 1994. Preferences for live saving programs: how the public discounts time and age. *Journal of Risk and Uncertainty* **8** 243-265.

Dasgupta, P., E. Maskin. 2005. Uncertainty and hyperbolic discounting. *American Economic Review* **95** 1290-1299.

DellaVigna, S., U. Malmendier. 2006. Paying not to go to the gym. *American Economic Review* **96 (3)** 694-719.

Delquié, P. 1997. 'Bi-matching': A new preference assessment method to reduce compatibility effects. *Management Science* **43** 640-658.

Diamond, P., B. Köszegi. 2003. Quasi-hyperbolic discounting and retirement. *Journal of Public Economics* **87** 1839-1872.

Dolan, P. 1997. Modeling valuations for EuroQol health states. *Medical Care* **35** 1095-1108.

Dolan, P. 2000. The measurement of health-related quality of life for use in resource allocation decisions in health care. A.J. Culyer, J.P. Newhouse, eds. *Handbook of Health Economics (vol. 1).* Elsevier, Amsterdam, 1723-1760.

Dolan, P., C. Gudex. 1995. Time preference, duration and health state valuations. *Health Economics* **4 (4)** 289-299.

Dolan, P., M.W. Jones-Lee. 1997. The time trade-off: a note on the effect of lifetime reallocation of consumption and discounting. *Journal of Health Economics* **16** 731-739.

Dolan, P., P.F.M. Stalmeier. 2003. The validity of time trade-off values in calculating QALYs: Constant proportional time trade-off versus the proportional heuristic. *Journal of Health Economics* **22** 445-458.

Donaldson, C., R. Thomas, D.J. Torgerson. 1997. Validity of open ended and payment scale approaches to eliciting willingness to pay. *Applied Economics* **29** 79-84.

Donkers, B., R. Gregory, J.L. Knetsch. 2004. Discounting future gains and future losses: predictable differences and appropriate measures. *Working paper.*

Dyer, J. S., R.K. Sarin. 1982. Relative risk aversion. *Management Science* **28** 875-886.

Etchart-Vincent, N. 2004. Is probability weighting sensitive to the magnitude of consequences? An experimental investigation on losses. *Journal of Risk and Uncertainty* **28** 217-235.

Fennema, H., M. van Assen. 1999. Measuring the utility of losses by means of the trade-off method. *Journal of Risk and Uncertainty* **17** 277-295.

Fischer, G.W., Z. Carmon, D. Ariely, G. Zauberman. 1999. Goal-based construction of preferences: Task goals and the prominence effect. *Management Science* **45** 1057-1075.

Fishburn, P.C. 1989. Retrospective on the utility theory of von Neumann and Morgenstern. *Journal of Risk and Uncertainty* **2** 127-158.

Fishburn, P.C., G.A. Kochenberger. 1979. Two-piece von Neumann Morgenstern utility functions. *Decision Sciences* **10** 503-518.

Fisher, I. 1930. *The theory of interest: As determined by impatience to spend income and opportunity to invest it*. Kelley and Millman, New York.

Frank, R.H., R.M. Hutchens. 1993. Wages, seniority, and the demand for rising consumption. *Journal of Economic Behavior and Organization* **21** 251-276.

Frederick, S. 1999. *Discounting, time preference, and identity.* Ph.D. thesis, Carnegie Mellon University.

Frederick, S., G.F. Loewenstein, T. O'Donoghue. 2002. Time discounting and time preference: a critical review. *Journal of Economic Literature* **40** 351-401.

Fuchs, V.R. 1982. Time preference and health: An exploratory study. V.R. Fuchs, ed. *Economic Aspects of Health.* University of Chicago Press, Chicago, 93-120.

Gafni, A. 1995. Time in health: can we measure individuals' 'pure time preferences'? *Medical Decision Making* **15 (1)** 31-37.

Gigliotti, G., B. Sopher. 2003. Analysis of intertemporal choice: a new framework and experimental results. *Theory and Decision* **55 (3)** 209-233.

Gravelle, H., W.B.F. Brouwer, L.W. Niessen, M.J. Postma, F.F.H. Rutten. 2007. Discounting in economic evaluations: Stepping forward towards optimal decision rules. *Health Economics* **16 (3)** 307-317.

Green, L., N. Fristoe, J. Myerson. 1994. Temporal discounting and preference reversals in choice between delayed outcomes. *Psychonomic Bulletin and Review* **1** 383-389.

Green, L., J. Myerson, E.W. Macaux. 2005. Temporal discounting when the choice is between two delayed rewards. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **31 (5)** 1121-1133.

Gruber, J.H., B. Köszegi. 2001. Is addiction "rational"? Theory and evidence. *Quarterly Journal of Economics* **116** 1261-1303.

Gruber, J.H., S. Mullainathan. 2005. Do cigarette taxes make smokers happier? *Advances in Economic Analysis & Policy* **5 (1)** Article 4.

Hall, J., K. Gerard, G. Salkeld, J. Richardson. 1992. A cost utility analysis of mammography screening in Australia. *Social Science & Medicine* **34 (9)** 993-1004.

Harris, C., D. Laibson. 2001. Dynamic choices of hyperbolic consumers. *Econometrica* **69** 935-957.

Harrison, G.W., M.I. Lau, M.B. Williams. 2002. Estimating individual discount rates in Denmark: A field experiment. *American Economic Review* **92 (5)** 1606-1617.

Harvey, C.M. 1986. Value functions for infinite-period planning. *Management Science* **32** 1123-1139.

Harvey, C.M. 1994. The reasonableness of non-constant discounting. *Journal of Public Economics* **53** 31-51.

Harvey, C.M. 1995. Proportional discounting of future costs and benefits. *Mathematics of Operations Research* **20** 381-399.

Herrnstein, R.J. 1981. Self-control as response strength. C.M. Bradshaw, E. Szabadi, C.F. Lowe, eds. *Quantification of steady-state operant behavior.* Elsevier, Amsterdam, 3-20.

Hertwig, R., A. Ortmann. 2001. Experimental practices in economics: A challenge for psychologists? *Behavioral and Brain Sciences* **24** 383-403.

Hesketh, B. 2000. Time perspective in career-related choices: applications of time-discounting principles. *Journal of Vocational Behavior* **57** 62-84.

Holcomb, J.H., P.S. Nelson. 1992. Another experimental look at individual time preference. *Rationality and Society* **4 (2)** 199-220.

Horowitz, J.K., K.E. McConnell. 2002. A review of WTA/WTP studies. *Journal of Environmental Economics and Management* **44** 426-447.

Hsee, C.K., R.P. Abelson, P. Salovey. 1991. The relative weighting of position and velocity in satisfaction. *Psychological Science* **2 (4)** 263-266.

Kahneman, D., J.L. Knetsch, R.H. Thaler. 1990. Experimental tests of the endowment effect and the Coase Theorem. *Journal of Political Economics* **98** 1325-1348.

Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47** 263-91.

Karp, L. 2005. Global warming and hyperbolic discounting. *Journal of Public Economics* **89** 261-282.

Keller, L.R., E. Strazzera. 2002. Examining predictive accuracy among discounting models. *Journal of Risk and Uncertainty* **24 (2)** 143-160.

Kirby, K.N. 1997. Bidding on the future: evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General* **126** 54-70.

Kirby, K.N., R.J. Herrnstein. 1995. Preference reversals due to myopic discounting of delayed reward. *Psychological Science* **6 (2)** 83-89.

Kirby, K.N., N.N. Marakovic. 1995. Modelling myopic decisions: evidence for hyperbolic delay-discounting within subjects and amounts. *Organizational Behavior and Human Decision Processes* **64** 22-30.

Kirby, K.N., N.M. Petry, W.K. Bickel. 1999. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General* **128** 78-87.

Kirby, K.N., M. Santiesteban. 2003. Concave utility, transaction costs, and risk in measuring discounting of delayed rewards. *Journal of Experimental Psychology: Learning, memory and cognition* **29 (1)** 66-79.

Kirsch, J., A. McGuire. 2000. Establishing health state valuations for disease specific states: An example from heart disease. *Health Economics* **9 (2)** 149-158.

Knetsch, J.L. 1989. The endowment effect and evidence of nonreversible indifference curves. *American Economic Review* **79 (5)** 1277-1284.

Koopmans, T.C. 1960. Stationary ordinal utility and impatience. *Econometrica* **28** 287-309.

Krantz, D.H., R.D. Luce, P. Suppes, A. Tversky. 1971. *Foundations of measurement (vol. 1)*. Academic Press, New York.

Krusell, P., A.A. Smith. 2003. Consumption-savings decisions with quasi-geometric discounting. *Econometrica* **71** 365-375.

Laibson, D. 1997. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* **112** 443-477.

Lamers, L.M., P.F.M. Stalmeier, P.F.M. Krabbe, J.J. Busschbach. 2006. Inconsistencies in TTO and VAS values for EQ-5D health states. *Medical Decision Making* **26 (2)** 173-181.

Lamers, L.M., P.F.M. Stalmeier, J. McDonnell, P.F.M. Krabbe, J.J. Busschbach. 2005. Kwaliteit van leven meten in economische evaluaties: Het Nederlands EQ-5D tarief. (In Dutch.) *Nederlands Tijdschrift voor Geneeskunde* **149 (28)** 1574-1578.

Lazaro, A., R. Barberan, E. Rubio. 2001. Private and social time preferences for health and money: An empirical estimation. *Health Economics* **10** 351-356.

LeBoeuf, R.A. 2006. Discount rates for time versus dates: The sensitivity of discounting to time-interval description. *Journal of Marketing Research* **43 (1)** 59-72.

Llewellyn-Thomas, H., H.J. Sutherland, R. Tibshirani, A. Ciampi, J.E. Till, N.F. Boyd. 1982. The measurement of patients' values in medicine. *Medical Decision Making* **2** 449-462.

Loewenstein, G.F. 1988. Frames of mind in intertemporal choice. *Management Science* **34 (2)** 200-214.

Loewenstein, G.F., D. Prelec. 1991. Negative time preference. *American Economic Review* **81** 347-352.

Loewenstein, G.F., D. Prelec. 1992. Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics* **107** 573-597.

Loewenstein, G.F., D. Prelec. 1993. Preferences for sequences of outcomes. *Psychological Review* **100 (1)** 91-108.

Loewenstein, G.F., N. Sicherman. 1991. Do workers prefer increasing wage profiles? *Journal of Labor Economics* **9** 67-84.

Loewenstein, G.F., R.H. Thaler. 1989. Anomalies: intertemporal choice. *Journal of Economic Perspectives* **3 (4)** 181-193.

Luce, R.D., H. Raiffa. 1957. *Games and decisions*. Wiley, New York.

MacKeigan, L.D., A. Gafni, B.J. O'Brien. 2003. Double discounting of QALYs. *Health Economics* **12** 165-169.

MacKeigan, L.D., L.N. Larson, J.R. Draugalis, J.L. Bootman, L.R. Bruns. 1993. Time preference for health gains versus health losses. *PharmacoEconomics* **3** 374-386.

Madden, G.J., W.K. Bickel, E.A. Jacobs. 1999. Discounting of delayed rewards in opioid-dependent outpatients: exponential or hyperbolic discounting functions? *Experimental and Clinical Psychopharmacology* **7** 284-293.

Martin, A.J., P.P. Glasziou, R.J. Simes, T. Lumley. 2000. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. *International Journal of Technology Assessment in Health Care* **16 (1)** 137-147.

Mazur, J.E. 1987. An adjusting procedure for studying delayed reinforcement. J.E. Mazur, M.L. Commons, J.A. Nevin, H. Rachlin, eds. *Quantitative analyses of behavior (vol. 5): The effect of delay and of intervening events on reinforcement value.* Erlbaum, Hillsdale, NJ, 55-73.

McCord, M.R., R. de Neufville. 1984. Utility dependence on probability: An empirical demonstration. *Journal of Large Scale Systems* **6** 91-103.

McNeil, B.J., R. Weichselbaum, S.G. Pauker. 1978. Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine* **299** 1397-1401.

McNeil, B.J., R. Weichselbaum, S.G. Pauker. 1981. Speech and survival: trade-offs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine* **305** 982-987.

Miyamoto, J.M., S.A. Eraker. 1988. A multiplicative model of the utility of survival duration and health quality. *Journal of Experimental Psychology: General* **117** 3-20.

Moore, M.J., W.K. Viscusi. 1990. Models for estimating discount rates for long-term health risks using labor market data. *Journal of Risk and Uncertainty* **3 (3)** 381-401.

Myerson, J., L. Green. 1995. Discounting of delayed rewards: models of intertemporal choice. *Journal of the Experimental Analysis of Behavior* **64 (3)** 263-276.

Nord, E. 1992. Methods for quality adjustment of life years. *Social Science and Medicine* **34 (5)** 559-569.

O'Donoghue, T., M. Rabin. 1999. Doing it now or later. *American Economic Review* **89** 103-124.

O'Donoghue, T., M. Rabin. 2006. Optimal sin taxes. *Journal of Public Economics* **90 (10-11)** 1825-1849.

Oliver, A.J. 2003. The internal consistency of the standard gamble: Tests after adjusting for prospect theory. *Journal of Health Economics* **22** 659-674.

Olsen, J.A. 1993. On what basis should health be discounted? *Journal of Health Economics* **12** 39-54.

Olson, M., M.J. Bailey. 1981. Positive time preference. *Journal of Political Economy* **89 (1)** 1-25.

Pennings, J.M.E., A. Smidts. 2000. Assessing the construct validity of risk attitude. *Management Science* **46** 1337-1348.

Pennings, J.M.E., A. Smidts. 2003. The shape of utility functions and organizational behavior, *Management Science* **49** 1251-1263.

Phelps, E.S., R.A. Pollak. 1968. On second-best national saving and game-equilibrium growth. *The Review of Economic Studies* **35** 185-199.

Pliskin, J.S., D.S. Shepard, M.C. Weinstein. 1980. Utility functions for life years and health status. *Operations Research* **28** 206-224.

Prelec, D. 2004. Decreasing impatience: A criterion for non-stationary time preference and 'hyperbolic' discounting. *Scandinavian Journal of Economics* **106** 511-532.

Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* **3** 323-343.

Rabin, M. 2000. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica* **68** 1281-1292.

Rachlin, H., A. Raineri, D. Cross. 1991. Subjective probability and delay. *Journal of the Experimental Analysis of Behavior* **55** 233-244.

Ramsey, F.P. 1928. A mathematical theory of saving. *The Economic Journal* **38** 543-559.

Read, D. 2001. Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty* **23 (1)** 5-32.

Read, D. 2004. Intertemporal choice. D. Koehler, N. Harvey, eds. *Blackwell handbook of judgment and decision making*. Blackwell, Oxford, 424-443.

Read, D., M. Airoldi, G. Loewe. 2005a. Intertemporal tradeoffs priced in interest rates and amounts: A study of method variance. *Working paper.*

Read, D., S. Frederick, B. Orsel, J. Rahman. 2005b. Four score and seven years from now: The date/delay effect in temporal discounting. *Management Science* **51 (9)** 1326-1335.

Redelmeier, D.A., D.N. Heller. 1993. Time preferences in medical decision making and cost-effectiveness analysis. *Medical Decision Making* **13** 212-217.

Rohde, K.I.M. 2005. The hyperbolic factor: A measure of decreasing impatience. *Working Paper.*

Rubinstein, A. 2003. 'Economics and psychology'? The case of hyperbolic discounting. *International Economic Review* **44** 1207-1216.

Rutten-van Mölken, M.P., C.H. Bakker, E.K.A. van Doorslaer, S. van der Linden. 1995. Methodological issues of patient utility measurement. Experience from two clinical trials. *Medical Care* **33** 922-937.

Sackett, D.L., G.W. Torrance. 1978. The utility of different health states as perceived by the general public. *Journal of Chronic Disease* **31** 697-704.

Salanié, F., N. Treich. 2006. Over-savings and hyperbolic discounting. *European Economic Review* **50** 1557-1570.

Samuelson, P.A. 1937. A note on measurement of utility. *The Review of Economic Studies* **4** 155-161.

Savage, L.J. 1954. *The foundations of statistics*. Wiley, New York.

Sayman, S., A. Onculer. 2006. Reverse time inconsistency. *Working paper.*

Schunk, D., C. Betsch. 2006. Explaining heterogeneity in utility functions by individual differences in decision modes. *Journal of Economic Psychology* **27** 386-401.

Shelley, M.K. 1993. Outcome signs, question frames and discount rates. *Management Science* **39** 806-815.

Shelley, M.K., T.C. Omer. 1996. Intertemporal framing issues in management compensation. *Organizational Behavior and Human Decision Processes* **66 (1)** 42-58.

Sopher, B., A. Sheth. 2006. A deeper look at hyperbolic discounting. *Theory and Decision* **60** 219-255.

Spencer, A. 2003. The TTO method and procedural invariance. *Health Economics* **12 (8)** 655-668.

Stalmeier, P.F.M., T.G.G. Bezembinder. 1999. The discrepancy between risky and riskless utilities: A matter of framing? *Medical Decision Making* **19** 435-447.

Stalmeier, P.F.M., T.G.G. Bezembinder, I.J. Unic. 1996. Proportional heuristics in time tradeoff and conjoint measurement. *Medical Decision Making* **16** 36-44.

Stalmeier, P.F.M., G.B. Chapman, A.G.E.M. de Boer, J.J.B. van Lanschot. 2001. A fallacy of the multiplicative QALY model for low-quality weights in students and patients judging hypothetical health states. *International Journal of Technology Assessment in Health Care* **17 (4)** 488-496.

Stalmeier, P.F.M., P.P. Wakker, T.G.G. Bezembinder. 1997. Preference reversals: Violations of unidimensional procedure invariance. *Journal of Experimental Psychology: Human Perception and Performance* **23** 1196-1205.

Starmer, C. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* **28** 332-382.

Shin, A.Y., P.J. Porter, M.C. Wallace, G. Naglie. 1997. Quality of life of stroke in younger individuals. Utility assessment in patients with arteriovenous malformations. *Stroke* **28** 2395-2399.

Stiggelbout, A.M., J.C.J.M. de Haes. 2001. Patient preference for cancer therapy: An overview of measurement approaches. *Journal of Clinical Oncology* **19 (1)** 220-230.

Stiggelbout, A.M., G.M. Kiebert, J. Kievit, J.W.H. Leer, J.D.F. Habbema, J.C.J.M. de Haes. 1995. The "utility" of the time trade-off method in cancer patients: Feasibility and proportional trade-off. *Journal of Clinical Epidemiology* **48 (10)** 1207-1214.

Stiggelbout, A.M., G.M. Kiebert, J. Kievit, J.W.H. Leer, G. Stoter, J.C.J.M. de Haes. 1994. Utility assessment in cancer patients: adjustment of time trade-off scores for the utility of life years and comparison with standard gamble scores. *Medical Decision Making* **14** 82-90.

Strotz, R.H. 1956. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies* **23** 165-180.

Thaler, R.H. 1981. Some empirical evidence of dynamic inconsistency. *Economics Letters* **8** 201-207.

Thaler, R.H., S. Benartzi. 2004. Save more tomorrow: Using behavioral economics to increase employee saving. *Quarterly Journal of Economics* **112** 164-187.

Tversky, A. 1967. Additivity, utility, and subjective probability. *Journal of Mathematical Psychology* **4** 175-201.

Tversky, A., D. Kahneman. 1986. Rational choice and the framing of decisions. *The Journal of Business* **59 (4)** S251-S278.

Tversky, A., D. Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* **56** 1039-1061.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5** 297-323.

Tversky, A., S. Sattath, P. Slovic. 1988. Contingent weighting in judgment and choice. *Psychological Review* **95** 371-384.

Unic, I., P.F.M. Stalmeier, L.C.G. Verhoef, W.A.J. van Daal. 1998. Assessment of the time-tradeoff values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. *Medical Decision Making* **18** 268-277.

van der Pol, M.M., J.A. Cairns. 2002. A comparison of the discounted utility model and hyperbolic discounting models in the case of social and private intertemporal preferences for health. *Journal of Economic Behavior and Organization* **49** 79-96.

van der Pol, M.M., L. Roux. 2005. Time preference bias in time trade-off. *European Journal of Health Economics* **6** 107-111.

van Nooten F., W.B.F. Brouwer. 2004. The influence of subjective expectations about length and quality of life on TTO. *Health Economics* **13 (8)** 819-823.

van Osch, S.M.C., W.B. van den Hout, A.M. Stiggelbout. 2006. Exploring the reference point in prospect theory: Gambles for length of life. *Medical Decision Making* **26** 338-346.

van Osch, S.M.C., P.P. Wakker, W.B. van den Hout, A.M. Stiggelbout. 2004. Correcting biases in standard gamble and time tradeoff utilities. *Medical Decision Making* **24** 511-517.

Varey, C.A., D. Kahneman. 1992. Experiences extended across time: evaluation of moments and episodes. *Journal of Behavioral Decision Making* **5 (3)** 169-185.

Vuchinich, R.E., C.A. Simpson. 1998. Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology* **6** **(3)** 292-305.

Wakker, P.P. 1994. Separating marginal utility and probabilistic risk aversion. *Theory and Decision* **36** 1-44.

Wakker, P.P. 2007. Explaining the characteristics of the power (CRRA) utility family. *Health Economics, forthcoming.*

Wakker, P.P., D. Deneffe. 1996. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science* **42** 1131-1150.

Wakker, P.P., A.M. Stiggelbout. 1995. Explaining distortions in utility elicitation through the rank-dependent model for risky choices. *Medical Decision Making* **15** 180-186.

Samenvatting

Economische baten worden vaak op verschillende momenten in de tijd ontvangen. Er zijn talrijke voorbeelden van economische toepassingen waarin de uitkomsten op verschillende momenten optreden. Dit zijn onder meer spaarbeslissingen van huishoudens, het milieubeleid van landen, investeringsbeslissingen van bedrijven, gezondheidsgerelateerde beslissingen van individuen, en scholingsactiviteiten van leerlingen.

In het merendeel van deze gevallen worden toekomstige uitkomsten minder gewaardeerd dan soortgelijke huidige uitkomsten, dat wil zeggen er is *positieve tijdsvoorkeur*. Hier zijn verschillende redenen voor. Een eerste reden is dat de toekomst bijna altijd wordt omgeven door onzekerheid, terwijl uitkomsten die onmiddellijk of in de meer nabije toekomst worden ontvangen, zekerder zijn. Dit leidt tot het disconteren van toekomstige uitkomsten.

Ten tweede is nut vaak concaaf in uitkomsten (afnemend marginaal nut). Dit betekent dat extra eenheden van een bepaalde uitkomst minder extra nut opleveren naarmate je al meer van die uitkomst hebt. Een tweede kopje koffie geeft bijvoorbeeld doorgaans minder nut dan het eerste kopje. Aangezien welvaart over de tijd stijgt door bijvoorbeeld economische groei, hebben mensen in de toekomst meer mogelijkheden om te consumeren. Het nut van deze extra consumptie stijgt vanwege het afnemende marginale nut echter niet evenredig met de consumptiestijging, zodat toekomstige uitkomsten minder nut geven dan dezelfde uitkomsten die nu verkregen worden, met als gevolg dat die toekomstige uitkomsten gedisconteerd worden.

Ten derde zijn mensen vaak kortzichtig en beschouwen zij niet altijd alle beschikbare informatie over de toekomst, hetgeen gelijkwaardig is aan het geven van een lager (of zelfs helemaal geen) gewicht aan toekomstige uitkomsten.

Het meten van tijdsvoorkeur

Tijdsvoorkeur heeft grote invloed op vele economische activiteiten. Het is daarom noodzakelijk om goede metingen van tijdsvoorkeur te verkrijgen. In verschillende wetenschappelijke disciplines, waaronder economie, psychologie en gezondheidswetenschappen, is er een debat gaande over de juiste manier om toekomstige baten te disconteren (bijv. Frederick e.a., 2002). Een aanzienlijk deel van de literatuur veronderstelt additiviteit van nut op verschillende tijdstippen. Dit betekent dat het totale gedisconteerde nut van alle baten kan worden verkregen door het vermenigvuldigen van de hoeveelheid nut in elke periode met een discontofactor, en het vervolgens optellen van deze gedisconteerde hoeveelheden nut. Dit impliceert dat het nut van een hoeveelheid baten op een bepaald punt in de tijd onafhankelijk is van de hoeveelheid baten op een ander punt in de tijd. Het meest gebruikte gedisconteerde nutsmodel is het constante disconteermodel, waarin de disconteerfunctie bepaald wordt door een constante jaarlijkse discontovoet. De praktijk van het disconteren van toekomstige nutsstromen door middel van een constante voet wordt echter betwist vanwege empirisch waargenomen schendingen van het constante disconteermodel (bijv. Ainslie, 1975; Thaler, 1981; Benzion e.a., 1989).

Hyperbolische disconteermodellen (bijv. Harvey, 1986; Loewenstein en Prelec, 1992) zijn populaire alternatieven. In deze modellen is de discontovoet niet constant maar is per periode verschillend. Baten die na een korte tijd optreden, krijgen een hoger discontopercentage per periode dan baten die na een langere tijd optreden. Met andere woorden, hyperbolisch disconterende mensen handelen alsof ze geduldiger worden naarmate uitbetalingen verder in de toekomst optreden. Er is daarnaast een aantal andere schendingen van het constante disconteermodel geobserveerd, zoals het op een verschillende manier disconteren van winsten en verliezen (Thaler, 1981; Loewenstein, 1988).

Een bezwaar tegen de meeste van de bestaande empirische studies over tijdsvoorkeur is dat deze studies lineair nut van geld veronderstelden, of dat zij veronderstelden dat de nutsfunctie een bepaalde parametrische vorm had. De daaruit voortvloeiende tijdsvoorkeurschattingen zijn derhalve onzuiver als deze veronderstelling niet geldig is. Een belangrijk doel van dit proefschrift was om dit probleem te verhelpen door middel van het aandragen en toetsen van nieuwe methodes om tijdsvoorkeur te meten die deze veronderstellingen niet hoeven te maken.

Daartoe heeft Hoofdstuk 2 eerst een overzicht van de bestaande literatuur over tijdsvoorkeur gegeven en laten zien dat de schendingen van constante discontering talrijk en systematisch zijn. In dat hoofdstuk zijn eveneens de implicaties van deze schendingen voor medische besluitvorming bediscussieerd en zijn diverse voorbeelden van toepassingen op gezondheidsgerelateerd gedrag gepresenteerd. Hierin is duidelijk gemaakt dat de schendingen van constante discontering te verklaren zijn met behulp van nieuwe modellen. Die nieuwe modellen kan men benutten om beleidsaanbevelingen te verbeteren.

Daarna is er, in Hoofdstuk 3, een methode geïntroduceerd die nut kan meten zonder een bepaalde parametrische vorm te veronderstellen en die vervolgens kan worden gebruikt om gemeten tijdsvoorkeur te corrigeren voor nutskromming. Bovendien meet de methode nut in een intertemporele context, zodat er gedurende het gehele meetproces een uniforme context gebruikt wordt. De methode is getoetst door middel van een experiment. De resultaten stelden mij in staat om de rol van nutskromming in het meten van tijdsvoorkeur te analyseren, om verschillende disconteermodellen met elkaar te vergelijken en om aan te geven welk model de data het beste beschreef. Het vergelijken van de fit van verschillende disconteermodellen was namelijk nog nauwelijks gedaan in eerdere studies. Ook heb ik getoetst of de asymmetrie tussen het disconteren van winsten en verliezen verklaard kon worden met behulp van verschillende nutsfuncties voor winsten en verliezen. Geconcludeerd werd dat intertemporeel nut concaaf was voor winsten en convex voor verliezen, consistent met een hypothese van Loewenstein en Prelec (1992). Nutskromming had echter weinig invloed op tijdsvoorkeur. Wel verminderde het de winstverliesasymmetrie enigszins, maar het verschil in discontovoeten tussen winsten en verliezen bleef significant. Een ander interessant resultaat was dat deze asymmetrie gevonden werd ondanks het gebruik van een neutraal kader. Daarom verwierpen de resultaten van dit hoofdstuk de stelling van Shelley (1993) dat de winstverliesasymmetrie kan worden verklaard door een kadereffect.

Voorts vond ik dat het hyperbolische disconteermodel van Loewenstein en Prelec (1992) de data significant beter beschreef dan het constante disconteermodel en andere hyperbolische disconteermodellen, zoals het quasi-hyperbolische disconteermodel. Dit laatste model is zeer populair in economische toepassingen en veronderstelt dat er sprake is van een *immediacy effect*. Dit betekent dat mensen een sterk onderscheid maken tussen baten die onmiddellijk verkregen worden en baten die in de toekomst verkregen worden, maar minder onderscheid maken tussen baten in de nabije toekomst en baten in de verdere toekomst. De discontovoet is met andere woorden hoog als onmiddellijke baten vergeleken worden met baten in de toekomst, maar is lager en constant als baten op verschillende punten in de toekomst vergeleken worden. Een implicatie van de resultaten van Hoofdstuk 3 is echter dat discontovoeten op een monotone wijze over de tijd dalen en, derhalve, dat hyperbolisch disconteren niet slechts door een *immediacy effect* wordt veroorzaakt zoals bij quasi-hyperbolisch disconteren.

In Hoofdstuk 4 is een andere methode geïntroduceerd die het mogelijk maakt om te toetsen of individuen afwijken van het constante disconteermodel en om hun afwijking van dit model te kwantificeren zonder de nutsfunctie te hoeven meten. De methode kan alternatieve tijdsvoorkeurmodellen toetsen. Zij maakt het door middel van een paar eenvoudige vragen mogelijk om te toetsen of individuen vatbaar zijn voor intertemporele arbitrage. Met behulp van een experiment is deze methode in de praktijk getoetst. In dit experiment zijn zowel schendingen van het constante disconteermodel als van het hyperbolische disconteermodel gevonden, terwijl de discontovoeten stegen bij grotere vertragingen, in tegenstelling tot de meeste bestaande gegevens en de resultaten van Hoofdstuk 3. De resultaten van Hoofdstukken 3 en 4 maken duidelijk dat geobserveerde tijdsvoorkeuren in sterke mate afhangen van de meetmethode. Het experiment van Hoofdstuk 3 gebruikte een keuzeprocedure, terwijl het experiment van Hoofdstuk 4 een *matching* procedure gebruikte. Bij een keuzeprocedure dient de respondent te kiezen tussen twee alternatieven en wordt er aan de hand van zijn antwoorden naar een indifferentiewaarde gezocht. Bij een *matching* procedure, daarentegen, wordt de respondent gevraagd om bij één alternatief een bepaalde waarde te geven die hem indifferent maakt tussen twee alternatieven. Daarnaast drukte Hoofdstuk 3 vertraging uit in maanden en jaren, terwijl Hoofdstuk 4 vertraging alleen in maanden uitdrukte. Een ander belangrijk verschil betreft de responsschaal. In Hoofdstuk 3 was dit geld, terwijl dit in Hoofdstuk 4 tijd was. Meer onderzoek naar de invloed van de procedure op de resultaten is daarom gerechtvaardigd.

In Hoofdstuk 5 is een nieuwe methode voorgesteld om tijdsvoorkeur voor toekomstige levensjaren te meten, ook wel bekend als het *nut van levensduur*. Het is belangrijk om kennis van deze nutsfunctie te hebben, omdat deze cruciaal is bij het doen van medische behandelingsaanbevelingen die het beste de belangen van de patiënt weergeven. De gebruikelijke manier om informatie over deze functie te verkrijgen is met behulp van de CE-methode, welke nut onder risico meet. Deze methode vereist dat de *verwachte nutstheorie*, de normatieve theorie voor beslissen onder risico, geldig is. Helaas is de verwachte nutstheorie niet zo goed in staat om de praktijk te verklaren (Starmer, 2000), zodat het via de CE-methode gemeten nut onzuiver kan zijn. Mensen hebben bijvoorbeeld vaak moeite om met kansen om te gaan, welke zij vaak een gewicht geven dat lager of hoger is dan de betreffende kans. Hier houdt de verwachte nutstheorie geen rekening mee. Bovendien heeft de CE-methode de uitkomst dood als stimulus. Dit leidt tot sterke risicoaversie en, daardoor, sterke concaafheid van nut (bijv. Tversky en Kahneman, 1986; Stiggelbout en De Haes, 2001; Bleichrodt e.a., 2003).

Het is daarom de moeite waard om nieuwe technieken te vinden om schattingen van de nutskromming van levensduur te verkrijgen die een risicovrije context gebruiken en het opnemen van de uitkomst dood vermijden. In dit proefschrift heb ik een dergelijke techniek voorgesteld, te weten de *risicovrije methode.* Deze methode is wederom in een experiment getoetst en vergeleken met twee bestaande methoden die het nut van levensduur onder risico meten (de CE-methode en de TO-methode). Daarnaast is er met behulp van een enquête onderzocht wat de respondenten van deze methoden vonden. De resultaten hiervan bevestigden dat zij de risicovrije methode gemakkelijker te beantwoorden vonden dan de andere twee methoden. Wat betreft de nutsfunctie voor levensduur leverde de CE-methode meer concaaf nut op dan de risicovrije methode. Dit verschil was echter niet langer significant na correctie voor kansweging. De resultaten van de TO-methode, die niet verstoord wordt door kansweging, verschilden niet significant van die van de risicovrije methode. Het lijkt er daarom op dat de risicovrije methode in staat is om een betrouwbare meting van nut te verschaffen en gemakkelijk toe te passen is voor praktische doeleinden.

Een andere opmerkelijke bevinding was dat nut beter door een exponentiële functie dan door de populaire machtsfunctie beschreven kon worden. Deze resultaten ondersteunen een constante *absolute* risicohouding met betrekking tot levensjaren in plaats van een constante *relatieve* risicohouding. Dit staat in tegenstelling tot andere studies die wel een constante relatieve risicohouding vonden. De precieze risicohouding van mensen heeft interessante gevolgen voor het *QALY*-model, een populair model om voorkeuren voor gezondheidsprofielen te beschrijven. Een constante absolute risicohouding, en de daarmee gepaard gaande exponentiële nutsfunctie voor levensduur, is consistent met constante discontering van toekomstige levensjaren. Een constante relatieve risicohouding, en de daarmee gepaard gaande machtsfunctie voor levensduur, is juist consistent met een hyperbolische discontering van toekomstige levensjaren.

Eén nutsbegrip

Dit proefschrift heeft ook bekeken of er één nutsbegrip bestaat dat geldig is in verschillende situaties of dat nut contextafhankelijk is en per domein verschilt. Economen betogen traditiegetrouw dat nut per domein verschilt en dat de nutsfunctie die relevant is voor besluitvorming onder risico daardoor niet kan worden toegepast in andere contexten, zoals besluitvorming onder zekerheid en intertemporele besluitvorming (zie Wakker, 1994, voor een overzicht). Op het gebied van de gezondheidseconomie is er echter een tendens om overdraagbaarheid van nut te veronderstellen. De TTO-methode, bijvoorbeeld, meet nut in een intertemporele context, maar de resulterende TTO-waarderingen worden dikwijls in economische evaluaties van de gezondheidszorg gebruikt, dat wil zeggen in welvaartseconomie. Hetzelfde geldt voor nut dat is gemeten met behulp van de *standard gamble* methode die beslissingen onder risico beschouwt.

Dit proefschrift heeft nutsfuncties voor geld en gezondheid in verschillende beslissingscontexten op een experimentele wijze gemeten. Een noviteit in dit proefschrift is dat nut voor geld is gemeten in een intertemporele situatie (Hoofdstuk 3). De resultaten zijn vergeleken met voorgaande nutsmetingen in een onzekere of risicovolle situatie. De bevindingen stemden in grote mate overeen, hetgeen een universeel nutsconcept ondersteunt. Daarnaast heeft Hoofdstuk 5 de risicovrije methode om het nut van levensduur in een risicovrije situatie te meten voorgesteld. De experimentele resultaten die met deze methoden werden verkregen, werden voor dezelfde respondenten vergeleken met de experimentele resultaten van twee andere bekende meetmethoden die een risicovolle situatie gebruiken. Er werden geen significante verschillen gevonden na correctie voor kansweging. Als we beseffen dat kansweging een afwijking is die niets met nutskromming te maken heeft, dan is deze bevinding opnieuw bewijs ten gunste van één nutsbegrip.

Deze resultaten hebben belangrijke implicaties. Zij steunen bijvoorbeeld de overdraagbaarheid van nut naar verschillende domeinen en daardoor de gebruikelijke praktijk in de gezondheidseconomie om TTO-nut (tijdsdomein) en *standard gamble* nut (risicodomein) te gebruiken in economische evaluaties (welvaartseconomiedomein). Bovendien verwerpen deze resultaten de gangbare opvatting in de economische wetenschap dat nut contextafhankelijk is.

De TTO-methode

Hoofdstukken 6, 7 en 8 pasten de resultaten toe op TTO, een belangrijke methode om het nut van gezondheid te meten. De meting van het nut van levensduur is gebruikt om de TTO-methode te corrigeren voor nutskromming. In een TTO dienen individuen een afweging te maken tussen kwaliteit van leven en duur van leven. Een probleem van de TTO-methode is echter de veronderstelling van lineair nut van levensduur, terwijl die vaak concaaf blijkt te zijn doordat veel mensen toekomstige levenstijd disconteren. Dit resulteert in een te laag nut voor gezondheidstoestanden (Bleichrodt, 2002). Het is wenselijk om dit verschil te kwantificeren en hiervoor te corrigeren.

Er zijn enkele eerdere pogingen ondernomen om TTO-scores te corrigeren voor het nut van levensduur (bijv. Stiggelbout e.a., 1994; Stalmeier e.a., 1996; Van Osch e.a., 2004; Van der Pol en Roux, 2005), maar de meeste van deze studies gebruikten de CE methode, wat vereist dat de verwachte nutstheorie geldig is. Als dit niet het geval is, zal de correctie van TTO-scores onjuist zijn. In dit proefschrift is de risicovrije methode aangewend om TTO-scores voor de nutskromming van levensduur te corrigeren, zodat men niet afhankelijk is van de geldigheid van de verwachte nutstheorie en de invloed van de uitkomst dood. De verschillen met ongecorrigeerde TTO-scores zijn onderzocht en de rol van nutscorrectie in diverse schendingen van de TTO-methode is verkend.

In Hoofdstuk 6 heb ik uitgelegd hoe de risicovrije methode gebruikt kan worden om TTO-scores voor nutskromming van levensduur te corrigeren en heb ik geschat hoe groot deze correctie is. Vanwege de concave vorm van de nutsfuncties waren de gecorrigeerde TTO-scores significant hoger dan de ongecorrigeerde TTO-scores. De grootte van dit verschil was ongeveer 0.05 (ongeveer 6%).

Hoofdstuk 7 toetste of TTO-scores afhankelijk zijn van de gebruikte meetprocedure. Als twee verschillende procedures om TTO te meten verschillende uitkomsten geven, kunnen de conclusies van economische evaluaties sterk afhangen van de gebruikte procedure. Het is daarom belangrijk te weten hoe deze verschillen veroorzaakt worden en welke procedure het beste gehanteerd kan worden. In Hoofdstuk 7 beschouwde ik de invloed van het nut van levensduur op de ongelijkheid tussen twee verschillende TTO-procedures. Mijn bevindingen waren dat corrigeren voor het nut van levensduur deze ongelijkheid vermindert, hoewel een groot en significant verschil overblijft. Dit wordt waarschijnlijk veroorzaakt door afkeer van verliezen, wat betekent dat mensen geneigd zijn om negatieve uitkomsten, bezien vanaf een bepaald referentiepunt, meer gewicht te geven dan soortgelijke positieve uitkomsten bezien vanaf dit referentiepunt (Tversky en Kahneman, 1991).

Ten slotte heeft Hoofdstuk 8 de CPTO-eigenschap van de TTO-methode behandeld. Deze eigenschap houdt in dat de gemeten TTO-score hetzelfde moet zijn voor verschillende duren. Als iemand bijvoorbeeld acht jaar in volledige gezondheid equivalent vindt aan tien jaar met rugklachten, dan zou hij acht maanden in volledige gezondheid equivalent moeten vinden aan tien maanden met rugklachten. Het bestaande bewijs over deze eigenschap is besproken en er is een nieuwe toets uitgevoerd om te onderzoeken of het nut van levensjaren op een constante proportionele manier werd afgeruild. Het bleek dat noch gewone levensjaren, noch het nut van levensjaren in een constante proportionele manier werden afgeruild. Dit resultaat is evidentie tegen de beschrijvende geldigheid van het QALY-model voor besluitvorming in de gezondheidszorg.

Dankwoord

Het feit dat mijn promotieonderzoek heeft geresulteerd in dit proefschrift is mede te danken aan de nodige mensen die nauw betrokken zijn geweest bij mijn onderzoek. Zonder anderen tekort te doen, wil ik er hier een aantal van noemen.

Allereerst wil ik mijn promotor Han Bleichrodt bedanken voor het begeleiden van mijn proefschrift gedurende de afgelopen vier jaar. Hij heeft mijn artikelen kritisch becommentarieerd, vele tips en ideeën gegeven en mij vrijgelaten mijn eigen onderzoekslijnen te ontwikkelen. Daarnaast bedank ik mijn tweede promotor Peter Wakker voor de privé-colleges, waarin hij alle geduld had om iets duidelijk uit te leggen en mijzelf het probleem liet oplossen. Bovendien heeft hij mij tijdens het schrijven van mijn scriptie enthousiast gemaakt voor zijn vakgebied, heeft hij diverse nuttige seminars en wetenschappelijke bijeenkomsten georganiseerd, en mij mede begeleid bij het schrijven van dit proefschrift.

De leden van de kleine commissie, Werner Brouwer, Daniel Read en Anne Stiggelbout, bedank ik voor het lezen en beoordelen van dit proefschrift, en voor het opponeren tijdens mijn verdediging.

Voorts ben ik dank verschuldigd aan de coauteurs van enkele hoofdstukken uit dit proefschrift. Naast Han en Peter zijn dit Mohammed Abdellaoui, Werner Brouwer en Kirsten Rohde. Het was prettig om met jullie samen te werken.

Ook zijn er mensen die mij, hoewel ze geen coauteur waren, hebben geholpen bij één of meerdere hoofdstukken. Deze personen ben ik allen zeer dankbaar. In het bijzonder heb ik veel hulp gehad van mijn broer Guido bij het programmeren van twee computergestuurde experimenten. Eveneens heb ik baat gehad van de nuttige adviezen die mijn collega's Tom, Stefan, Bas en Kirsten mij gaven over de opzet van één van die experimenten.

Ik ben het NWO dankbaar voor de toegekende VIDI-subsidie aan het project "The valuation of health," waar dit proefschrift deel van uitmaakt. Tevens bedank ik het College van Bestuur van de Erasmus Universiteit voor het nomineren van mijn voordracht voor de Chorafasprijs 2006, en ben ik de Dimitris N. Chorafas Foundation dankbaar voor het aan mij toekennen van deze prijs en de bijbehorende beurs, die kon worden gebruikt om een deel van dit proefschrift te financieren.

Tijdens mijn periode als Aio heb ik verschillende kamergenoten gehad met wie ik leuke tijden heb doorgebracht. Bij iBMG waren dit Martine, Martijn, Samir en Anne. Bij de FEW waren dit Tom en Kirsten. Met Tom heb ik het langste een kamer gedeeld. Wij leerden elkaar meer over de Belgische respectievelijk Nederlandse cultuur en maatschappij, en over het plaatselijke taalgebruik, hoewel het in ons eigen land toch altijd beter was dan in het andere. Bovendien gaf hij me de nodige statistische tips.

In addition, I would like to thank Louis Cabanilla for the great time we had during my first year in Rotterdam. I'm looking to forward to play an intensive tennis match with you again and to watch Oranje together with a good glass of beer.

Mijn paranimfen Guido en Joost bedank ik voor het mede organiseren van de promotieplechtigheid en alles wat hierbij hoort, en voor het mij terzijde staan tijdens de verdediging. Verder bedank ik mijn ouders Jaap en Anja voor de steun

gedurende mijn studie en periode als Aio, voor hun interesse in mijn promotie en in het bijzonder voor de goede verzorging toen ik een zware enkelfractuur had.

Natuurlijk bedank ik ook de leden van mijn korfbalvereniging RSKV Erasmus, waar ik een aantal prachtige jaren heb gehad en waar ik nog steeds actief ben. Ik kon er mijn gedachten verzetten, heb er veel vrienden gemaakt en veel geleerd op onder andere korfballend en bestuurlijk gebied. De volgende leden wil ik er hier even uitlichten: Jan Willem, Joep, Joost en Roy, de bakjes koffie, lunches en diners in de mensa met jullie zorgden voor de nodige ontspanning en vermaak.

About the author

Arthur Attema was born in Almelo on November 24, 1981, and grew up in Almere. After graduating from high school in 1999, he started studying General Economics at the University of Amsterdam. There he was student assistant at the Department of Quantitative Economics. In July 2003 he obtained his MSc degree (cum laude) after finishing a master thesis about decision making under risk, which was supervised by Peter Wakker. During the period of writing this thesis, he met Han Bleichrodt and agreed to start a PhD position under his supervision at the institute of Health Policy and Management (iBMG) of the Erasmus University Rotterdam.

He moved to Rotterdam where life turned out to be nice. In his first year as a PhD student he obtained an MSc degree in Health Economics. From the second year on, he was a staff member of the Health Economics/HEPL master course "Advanced Economic Evaluation", teaching on time preference and conjoint analysis. He also supervised master theses in the field of health economics and intertemporal choice.

In 2005, he joined Han Bleichrodt to the Faculty of Economics, where a new Health Economics Department was created. He participated in several conferences, including the FUR conferences in 2004 in Paris and in 2006 in Rome, the SPUDM conferences in 2005 in Stockholm and in 2007 in Warsaw, the Decision and Uncertainty Workshop in 2006 in Paris, and the QALYs Workshop in 2003 in Alicante. In the summer of 2006, he visited the GRID department of CNRS-ENSAM in Paris. Arthur's PhD research was awarded the Dimitris N. Chorafas Prize 2006, a prize for applied research.

After finishing his PhD thesis, Arthur was a Health Economics teacher at the Faculty of Life Sciences of the Free University in Amsterdam. At present, he works as a researcher at the National Institute for Public Health and the Environment (RIVM) in Bilthoven. His main activities at RIVM are the modeling of uncertainty in cost-effectiveness studies and studying the efficiency of several programs to reduce unhealthy behavior.

Arthur has been involved in several sportive activities. He was a member of the faculty indoor soccer team for four years, and he was captain and team manager of this team during more than three years. Moreover, he is an active korfball player. Since 2004 he is playing for RSKV Erasmus in Rotterdam. He is also treasurer of this korfball association and performs various voluntary tasks. Arthur further likes to play tennis, both at his tennis club in Almere and at the university campus, and likes cycling and running. Finally, he is a dedicated supporter of PSV Eindhoven and regularly visits matches of PSV in the Philips Stadium in Eindhoven, the most beautiful and attractive stadium of the Netherlands.